

ALFRED

Personal Interactive Assistant for Independent Living and Active Ageing



WP8– Piloting & Validation

D8.1.2: Piloting Definitions

Deliverable Lead: CHA

Contributing Partners: ESE, NFE

Delivery Date: 04/2015

Dissemination Level: Public

Version 1.0

The purpose of this deliverable is to present the first results of the pre-prototype evaluation. This was done by analysing the data that was collected in first Wizard of Oz testing sessions that were performed in France, the Netherlands and Germany. Furthermore additional specifications about the interrelated planning of the three pilots from a usability-, health-, user-, and economic perspective will be provided.



Document Status	
Deliverable Lead	Florian Feldwieser, CHA
Internal Reviewer 1	Tasos Martidis, TIE
Internal Reviewer 2	Emilia Garcia Nieto, WORLD
Type	Deliverable
Work Package	WP8 Piloting & Validation
ID	D8.1.2 Piloting Definitions
Due Date	04/2015
Delivery Date	31.03.2015
Status	For Approval

Document History	
Contributions	<p>V0.1, CHA, 29.10.2014;ESE, 15.12.14, CHA 16.12.14</p> <p>V0.2, NFE, 24.12.2014 evaluation and common framework, CHA 06.01.2015</p> <p>V0.3 12.01.2015 CHA, Ethical issues, Pilot in France: ESE, 21.01.2015</p> <p>V0.4, 26.01.2015 NFE</p> <p>V0.5,02.01.2015 NFE, 04.02.2015 CHA, 11.02.2015 ESE</p> <p>V0.6 +ESE+NFE+CHA 23.02.2015</p> <p>V0.7 26.02.2015 CHA</p> <p>V0.8 27.02.2015 NFE, 03.03.2015 CHA, 04.03.2015 ESE</p> <p>V0.9 04.03.2015 CHA</p>
Final Version	V1.0 30.03.2015

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.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 2 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

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

In this document we expose the usability tests that were performed in the first pre-prototype evaluation of ALFRED by using the Wizard of Oz (WoZ) method, where instead of a real system a human controls it. The WOZ technique is a method that simulated the speech recognition and dialogue capabilities of the app. It allows exploratory evaluation of an app before investing large amounts of development time in a product that in the end might not meet user demands. The missing functionality of the system is replaced by the “Wizard”. The wizard is a person from the ALFRED team that conducted the iterative evaluations. While the participants actually thought they were communicating with ALFRED, the intelligence behind the system was simulated by the wizard. The iterative sessions further were video- and/or audio recorded. These recordings helped with the further analysis after the testing sessions. The video and audio recording were reviewed by the test leaders after the test session and helped to uncover usability problems and other issues that occurred during the WOZ testing.

The main tested outcome was the interaction of the verbal communication between ALFRED test participants and the WoZ prototype which took place in M13 and M16 in France, the Netherlands and Germany. Overall six tasks with twelve older end users were tested and videotaped for in-depth analysis of the interaction after the test session. The tasks included testing the activation of ALFRED, setting an alarm, sending a message, calling someone, adding a reminder to the agenda and react to automatic suggestion given by ALFRED. The tasks were analysed with the After Scenario Questionnaire (ASQ) and the Post Study System Usability Questionnaire (PSSUQ) as well as Think Aloud methods and observation. Errors during the session were captured and proposed changes to resolve them are presented. Feedback from given from the participants was collected during the session and positive and negative remarks of the participants are specified.

Participants had to fill in three questions after completion of each task scenario in the ASQ and the PSSUQ. Rating options ranged from ‘I completely agree’ (value 1) to ‘I completely disagree’ with a total of seven options which were converted to a numerical value from 1 to 7 (value 7).

The ASQ and PSSUQ ratings overall were very positive and participants were very satisfied with the system and the idea of voice interaction. Most users experienced little or no problems when working with ALFRED. The main findings of for improvement however were, that participants want to be able to turn off ALFRED voice in some cases for privacy reasons, it was not always clear to see if ALFRED was already activated or not. Also participants were not totally satisfied with the voice of ALFRED and would prefer a less synthetic and slower speech output. Participants also forgot to mention specific dates and times when setting up an appointment. Despite the very positive ASQ and PSSUQ ratings it needs to be mentioned that the currently tested system used a simulated WoZ simulation and did not yet test the actual speech recognition abilities of ALFRED. Further iterative tests should be able to reveal if the actual ALFRED system will have similar speech recognition and dialogue capabilities.

The second half of this deliverable gives an overview of the further planning of the work in the ALFRED pilots. It explains the pilot methodology and four perspectives that will be used as a common framework to evaluate the results of ALFRED. The aim of the pilots is to test the real-world applicability of ALFRED in a holistic approach, in each pilot. The four

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 4 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

perspectives consist of the Usability Perspective, the Health Perspective, the Economic Perspective and the End User Perspective. For each perspective methods for evaluation are proposed as well as Key Performance Indicators (KPI's). These methods for evaluation and KPI's will be updated in D8.2.1 in M28 before the actual pilot running, based on the upcoming iterative testing results and the system prototype development. The final section describes for each separate pilot the planning, evaluation methods and practical implementation of the pilot. Although each pilot has its own characteristics, they will be able to use the common pilot methodology, thanks to the four perspectives.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 5 / 62
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	11
1.4	Abbreviations and Glossary	11
1.5	Document Structure	11
2	Iterative Evaluations	12
2.1	Summary	12
2.2	Methodology	12
2.3	Participants	13
2.4	Evaluation Tasks/Scenarios	15
2.5	Methods and Results	15
2.5.1	ASQ Ratings	16
2.5.2	PSSUQ Ratings	16
2.5.2.1	Overall Rating	17
2.5.2.2	System Quality	17
2.5.2.3	Information Quality	17
2.5.2.4	Internal Quality	17
2.5.2.5	Errors	17
2.5.3	Feedback from Participants	19
2.5.3.1	Positive Comments	19
2.5.3.2	Negative Comments	19
2.5.3.3	Recommendations	19
2.6	Conclusion	24
3	Pilot methodology	26
3.1	Methodology	26
3.1.1	Outline of the Pilot Perspectives	26
3.2	Usability Perspective	27
3.2.1	Envisaged Methods for Evaluation	28
3.2.2	Usability KPI's	28
3.2.3	Output and Upcoming Deliverables	28
3.3	Health perspective	28
3.3.1	Envisaged Methods of Measurements	29
3.3.2	Health KPIs	30
3.3.3	Output and upcoming deliverables	30
3.4	Economic perspective	30
3.4.1	Envisaged Methods of Evaluation	31
3.4.2	KPIs	31
3.4.3	Output and Upcoming Deliverables	32
3.5	End user Perspective	32
3.5.1	Envisaged Methods of Measurement	32
3.5.2	KPIs	34
3.5.3	Output and upcoming Deliverables	34
3.6	Overview Pilot Methodology	34
3.6.1	ALFRED Pillars and Pilots	34

3.6.2	KPI Overview	35
3.7	Practical Implementation of the Pilot Methodology	37
3.7.1	Time Plan	37
3.7.2	Recruitment.....	38
3.7.2.1	Profile Test Participants	38
3.7.2.2	Amount of Test Participants.....	38
3.7.2.3	Drop-out Strategy	39
3.7.3	Helpdesk.....	39
3.7.4	Workbook.....	40
3.7.5	Overview Equipment.....	40
3.7.6	Ethical management process.....	41
4	Individual Pilot Descriptions	42
4.1	Specifications for Pilot 1 – Individual Usability.....	42
4.1.1	Evaluation perspectives in Pilot 1	42
4.1.2	Involved Test Persons in Netherlands	43
4.1.3	Timeline and Pilot 1 Phases.....	44
4.1.4	Methodologies for Pilot 1.....	44
4.2	Specifications for Pilot 2 – Hospital Environment	45
4.2.1	Evaluation perspectives in Pilot 2	45
4.2.2	Involved Test Persons in Germany	46
4.2.3	Timeline for Pilot 2	46
4.2.4	Methodologies for Pilot 2.....	47
4.3	Specifications for Pilot 3 – Day-to-day usage	47
4.3.1	Evaluation Perspectives in the Pilot 3	47
4.3.2	Involved Test Persons in France.....	48
4.3.3	Timeline and the Pilot 3 Phases.....	49
4.3.4	Methodologies for Pilot 3.....	50
5	Conclusion	51
	References	52
	Annex A – ASQ Questionnaire	53
	Annex B – PSSUQ Questionnaire	55
	Annex C – ASQ Ratings	61
	Annex D – PSSUQ Ratings	62

List of Figures and Tables

List of Figures

Figure 1: Screenshot from the WOZ Frontend on a Laptop	16
Figure 2: Overall Framework of ALFRED Impact.....	26
Figure 3: Infographic – Overview of the Pilot Plan	27
Figure 4: Technology Acceptance Model [Davis, 1989].....	33

List of Tables

Table 1: Summary of the ALFRED Pre-Prototype Test Participants	13
Table 2: Background Variables of the Pre-prototype Test Participants.....	14
Table 3: Feedback Summary for Task 1	20
Table 4: Feedback Summary for Task 2	20
Table 5: Feedback Summary for Task 3.....	21
Table 6: Feedback Summary for Task 4	22
Table 7: Feedback Summary for Task 5.....	22
Table 8: Feedback Summary for Task 6	23
Table 9: General Feedback Summary	23
Table 10: KPI for the Usability Evaluation.....	28
Table 11: KPI for the Health Care Evaluation	30
Table 12: KPI for the Economic Evaluation.....	31
Table 13: KPI for the End User Evaluation	34
Table 14: Initial ALFRED Pilot Methodology Overview	35
Table 15: Overview of the KPIs Used in the Pilots.....	36
Table 16: Overview Piloting Time Plan	37
Table 17: Overview of Involved Target Groups in the Different Pilots.....	38
Table 18: Amount of Test Participants in the Pilot Countries	39
Table 19: Testing Material for the Pilot Phase	40
Table 20: Definition of the Involved Target Groups in Pilot 1	43
Table 21: Initial Time Planning for Pilot 1.....	44
Table 22: Initial Time Planning for Pilot 2.....	46
Table 23: Definition of the Involved Target Groups in the Pilot 3	48
Table 24: Initial Time Planning for Pilot 3.....	49

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.

The purpose of this deliverable is to present the first results of the pre-prototype evaluation. This was done by analysing the data that was collected in first Wizard of Oz testing sessions that were performed in France, the Netherlands and Germany. Furthermore additional specifications about the interrelated planning of the three pilots from a usability-, health-, user-, and economic perspective will be provided.

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 stuff 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,

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 9 / 62
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

and Human-Computer Interaction. For more information, please refer to the project website at <http://www.alfred.eu>.

1.2 Deliverable Purpose, Scope and Context

This deliverable will reflect the planning of the tasks in WP8. The pilot evaluations in WP8 consist of two different types of evaluations. First of all the pilot site partners ESE, NFE and CHA will organize evaluations on the early prototypes of ALFRED together with the end-user partners. This early prototype testing will prevent expensive development tasks that are not in line with end user expectations and requirements. These evaluations will be organized iteratively according to the outcome of different early prototypes and will focus on usefulness and usability for older persons. They will be organized in a controlled environment together with researchers and older people. These evaluations will be called in this deliverable; iterative evaluations. The results of the first iterative evaluations will be presented in this deliverable.

In the second part of the deliverable, the methodology of the pilot will be presented.. During these pilots the end users will test the ALFRED solution during a longer period of time in a normal situation in the home surrounding. The pilots will be organized within a framework of four perspectives. These perspectives will be introduced, together with evaluation methods and KPI's. This document will first describe the methodology that provides for a common framework for the pilots in Germany, France and the Netherlands.

Each pilot will perform more than just user interviews and trials. It will also create everything that is needed for the pilot. More precisely, this means that the pilots will create pilot specific apps in close collaboration with the app building tasks of the work packages, namely 4.5, 5.5, 6.5 and 7.4. While WP4-7 will concentrate on the core technical aspects of the app development (i.e., programming), the WP8 activities will concentrate on providing the pilot specific data and the logic as well as the app conceptualization and all other elements that go beyond the pure build and deployment process. Thanks to the iterative evaluations this process is implemented in an early stage and will run throughout the project.

Pilots of ALFRED will be performed in different EU member states. The first pilot will be performed in the Netherlands (T8.2), the second one in Germany (T8.3) and the third one in France (T8.4). It can, however, not be expected from elderly people to speak or understand proper English in call cases. As such, each prototype will start with a localization of the project results. This localization will include the translation of apps and component elements into different languages but it will also include the adaptation of other cultural environments such as units or time zones. Each pilot contains one or more technical partners which will support this adaptation process from a technical level. This work is already ongoing within the scope of the technical WPs and thanks to the iterative evaluations can be tackled in an early phase.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 10 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

1.3 Document Status and Target Audience

This document is listed in the Description of Work (DoW) as “public”, as it provides general information about the goals and scope of ALFRED and can therefore be used by external parties in order to get according insight into the project activities.

While the document primarily aims the project partners, this public deliverable can also be useful for the wider scientific and industrial community. This includes other publicly funded projects, which may be interested in collaboration activities.

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

This document describes the results of the iterative testing cycles and a test plan for conducting the Pilots. Chapter 2 describes the findings and results of the iterative testing sessions that were performed France, Germany and the Netherlands, and methodology that is employed in the testing procedure that will take place in. Chapter 3 gives information on the pilot methodology, and chapter 4 will provide information about the individual pilot descriptions details of the employed methodology that will be used to conduct the individual usability (Pilot 1), the hospital environment (Pilot 2) and the Day-to-day usage (Pilot 3). Chapter 5 gives a brief summary of the document.

The annex contains background documents that are needed for the testing procedure like, questionnaires, After scenario questionnaire (ASQ) and Post-study System Usability Questionnaire (PSSOQ) forms and test results.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 11 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

2 Iterative Evaluations

2.1 Summary

The first iterative evaluation of ALFRED took place in M13 and M16. This took place in The Netherlands, France and Germany, among twelve Test Persons (TP) in total, six Dutch, three German, two French and one Italian, nine of them female and three male, with an average age of 74 (SD = 5.74). The test consisted of six tasks and took approximately one hour per participant.

These first evaluations aimed to test the speech interaction with older end users. The evaluations helped to gather information about how the user speaks to ALFRED, in order to guide the design of ALFRED (core and apps). The tests help to gather recordings to enable evaluations of Automatic Speech Recognition (ASR). Additionally it enabled evaluation of Text-To-Speech synthesizer (TTS) and speaker quality by detecting the test participant's ability to hear and understand what ALFRED said. Furthermore, it enabled evaluation of the suitability of some selected utterances by detecting the TP's ability to understand what ALFRED said and meant.

As this was a first evaluation, the tests were performed using the Wizard of Oz technique (WoZ) instead of a real system beyond human control.

This report will consist of a description of these tests and the first results. First, an overview of the methodology and participants will be given. Then, the tasks will be described shortly, after which an extensive results section will follow, describing not only the results from the After scenario questionnaire (ASQ) and the Post-study System Usability Questionnaire (PSSUQ). The ASQ gives an indication which tasks are most problematic for the participants to execute. The PSSUQ helps to assess the user's perceived satisfaction with the system. A detailed description of these two tests can be found in D8.1 Piloting Definition in chapter 2.6.3. ASQ and the PSSUQ tests but also the errors made by the user. Furthermore, comments from the users (both positive and negative) will be addressed. Finally, an overview of recommendations based on the errors and comments will be given at the end of the report. The reader should keep in mind that these are the very first results only. More iterative evaluations will take place after M 18. Results of these tests will be reported in D8.2 around M28.

2.2 Methodology

TP's were recruited by the end user partners, using national projects and services to reach them. The sessions were organized in the offices of end user partners and were voice recorded with a recording application on the smartphone.

As mentioned the Wizard of Oz technique was applied with a "fake" application being used which is not controlled by user interaction but by someone else. The user talked to the smartphone but the reply was controlled by someone behind a computer, clicking the right responses (while the user was unaware of this, and believed to be interacting with the system itself).

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 12 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

Participants were provided a workbook with all necessary documents. They received information about the evaluation and had to sign an informed consent first. Next, they had to fill in a questionnaire regarding demographics. After they filled this in, they could start with the task. There were six tasks and every individual task was followed by the After Scenario Questionnaire (ASQ); a short three-question questionnaire measuring satisfaction with ease of use, time on task and help from the system.

After completing all tasks (or, trying as good as possible to complete them), there was one last questionnaire for participants to fill in. This was the Post-Study System Usability Questionnaire (PSSUQ), an 18-item questionnaire measuring overall satisfaction with the system, divided by four sub measures (overall satisfaction, system quality, information quality and interface quality). These questionnaires and how to use them have also been described thoroughly in D8.1. See Annex A for the ASQ and Annex B for the PSSUQ questionnaires. Annex C and Annex D provide the results of the ASQ and PSSUQ scores. The overall session lasted between 45 and 60 minutes per participant.

2.3 Participants

Evaluations were performed in the Netherlands, France and Germany with a total of twelve participants. Six of the participants were Dutch, two had French nationality, three were German and one was Italian. There were six female participants. The average age of the participants was 74 (SD = 6.69) with the youngest participant being 62 and the oldest 82 years old.

Table 1: Summary of the ALFRED Pre-Prototype Test Participants

	Netherlands (NFE)	Germany (CHA)	France (ESE)	Total
Number of participants	6	3	3	12
Number of Female and Male participants	4 (F), 2 (M)	3 (F)	2 (F), 1 (M)	9 (F), 3 (M)
Average age of the participants	78 years	71 years	70 years	73 years
Household income	1,666 €	2,500 €	n/a	2,083 €

Three of the participants were married, three were widowed, two were single and one was divorced. Of these participants, six lived on their own, two lived with their partner (but without children) and one participant filled in 'other'. All but one participant still lived independently. One participant lived in a congregate housing, so independent with the possibility to receive support at home.

When asked about their health, participants were quite positive about their own health. Nobody rated their health lower than average (four participants) and some rated themselves even as having a good health (three) or having a very good health (two). Self-rated vision and hearing was asked as well. For the former, seven participants reported no difficulties reading something and two reported having minor difficulties when trying to read something. Participants rated their own hearing ability to be slightly reduced as opposed to

their vision but still quite good, as six participants said to have no difficulties hearing and three indicated to only have minor difficulties. One's own motor abilities were rated positive as well, with seven participants stating to have no difficulties and two to have only minor difficulties. There was a bit more difference in the rating of one's own mobility level: Three participants rated this to be fair, three rated it to be good, two rated it to be excellent and one participant rated it to be "not so good, unfortunately".

Participants had a diverse level of education; four of them had a lower vocational education, two had intermediate vocational education, one had higher vocational education and two had a university degree. All participants were retired at the time of testing; five of them still did work as a volunteer.

The diversity in technological experience was high (but could be seen as bell-shaped curved), with three participants rating their own experience as average, two as low, two as high, one as very low and one as very high. Eight participants rated their attitude towards technology as positive, the last participant rated it to be neutral.

Table 2: Background Variables of the Pre-prototype Test Participants

Variables	Answering options (<i>numbers</i>) (N = 12)
Average Age (years)	74 ± 5.74
Gender	Male (3) , Female (9)
Nationality	Dutch (6) , German (3) , French (2) , Italian (1)
Marital status	Single (3) , Married (5) , Divorced (1) , Widowed (3)
Living situation	Living alone (7) , Living with children, Living with partner (4) , no children, Living with partner and children, other (1)
Education	Primary education, Lower vocational education (4) , Intermediate vocational education (2) , Intermediate secondary education, Higher secondary education (3) , Higher vocational education (1) , University (2)
Average income (€)	1,875 ± 649.52
Employment: <i>Multiple answers possible</i>)	Retired (11) , Employed, Unemployed, Voluntary work (5) Other (1)
Residential situation	Living independently at home (11) , Independent planned housing, congregate housing (1) , Home for older people
Self-rated Health	Poor, Not so good, Fair (4) , Good (5) , Very good (3)
Vision: <i>Is your eyesight good enough to read ordinary newspaper print? (with glasses if usually worn)</i>	Yes without difficulty (9) , Yes, with minor difficulty (3) , Yes, with major difficulty, No, not able to
Hearing: <i>Do you hear what is said in a conversation with 3 or 4 other persons? (with hearing aid if you wear one)</i>	Yes without difficulty (8) , Yes, with minor difficulty (4) , Yes, with major difficulty, No, not able to
Motor control: <i>Can you press small items like buttons on a remote control?</i>	Yes without difficulty (10) , Yes, with minor difficulty (2) , Yes, with major difficulty, No, not able to
Self-rated mobility level	Poor, Not so good (1) , Fair (3) , Good (4) , Very good (4)
Technological experience	Very high (2) , High (2) , Medium (4) , Low (2) , Very low (2)
Attitude towards technology	Positive (10) , Neutral (2) , Negative
Use of ICT tools	Daily Use (7) , Weekly Use (3) , Monthly Use, Rarely (1) , Never (1)
Type of phone: <i>Multiple answers possible</i>	Normal mobile phone (5) , Senior mobile phone (with big buttons) (2) , Smartphone (Android, e.g. Samsung) (5) , Smart phone (Apple), Other: Normal phone (1)
Online activities: <i>Multiple answers possible</i>	E-mail (4) , Chat (WhatsApp) (5) , Look up a recipe (2) , Web search (3) , Look for health/medical information (2) , Look for information on a hobby or interest (2) , Look for transport

	information (2), Look on a map for directions (3), News (2), Weather (3), Sports (1), Online banking (4), Shopping (2), Playing games (4), Job search, Financial/stock trading, Visit a local, state or federal government website (1), none (4) Other:
--	---

When asked how many times participants used ICT, six said they used it on a daily basis, two said they used it at least weekly and one claimed she never used ICT. As ALFRED is an application for a smartphone, the participants were also asked about their type of phone. One participant only had a normal phone (no mobile), three had a normal mobile phone, two had a senior mobile phone with extra-large buttons, and four users had a smartphone (Android type). The total amount of answers here was ten, as one user indicated to have two phones; a smartphone for everyday life and a simpler, cheaper normal mobile phone for vacation. Users varied in what applications they used. Mentioned the most were e-mail, WhatsApp, looking for information about health, recipes, one's hobby, the weather, public transport or traffic. Other applications mentioned were applications for playing games, reading the news, online banking and online shopping. A short overview of all variables about participants' background is also displayed in table 2 above.

2.4 Evaluation Tasks/Scenarios

The task scenarios used were derived from the user stories (see D8.1) which were written down after the initial focus group sessions. The scenarios used in these evaluations were based on what the ALFRED (WOZ) prototype was capable of and which of the scenarios were the most important for usage in everyday life when interacting with ALFRED. The scenarios chosen were the following:

- Activate ALFRED
- Set an alarm
- Send a message
- Call someone
- Add a reminder to your agenda
- React to an automatic suggestion given by ALFRED

2.5 Methods and Results

The evaluation was performed with two people guiding the session and one participant. One person guided the user through the questionnaires and questions, another was the "wizard" (WOZ) and monitored the user's utterances to the ALFRED system and responded in a proper fashion by means of a program on the laptop, which was connected to the smartphone (also see Figure 1 for a screenshot of this program). In one session from NFE, there was a third person who took pictures. All participants were able to complete all scenarios, albeit not in a perfect, direct way.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 15 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

Figure 1: Screenshot from the WOZ Frontend on a Laptop

2.5.1 ASQ Ratings

Participants had to fill in three questions after completion of each task scenario. Rating options ranged from 'I completely agree' to 'I completely disagree' with a total of seven options which were converted to a numerical value from 1 to 7. In this sense, lower values represent a higher satisfaction and vice versa, higher values represent a lower satisfaction.

In general, participants were very satisfied with the system. Except for two, all average values on all three levels for all six scenarios were below 2, with all standard deviations for these values also ranging below 2. The highest value was 2.17 (SD = 1.97), for the help of the system given and ease of completing by the first task. Even though this value is far in the range of a high satisfaction score, the lower value here can be explained by the fact that in the first task, ALFRED has to be activated and thus, the system does not give any feedback yet before it is activated. This task required some knowledge on how to handle modern smartphones, such as unlocking a screensaver or activating apps on a touchscreen display. Since no information was provided how these operations should be performed users with low or very low technological experience (n=4) had difficulties performing this task. Lowest average values were seen for time on task when calling someone (M = 1.11, SD = 0.33) the fourth scenario tested) and ease of use for adding a reminder to the agenda (M = 1.22, SD = 0.44).

It must be noted that scores for the category 'satisfaction with help from the system' are slightly flawed. Some of the Dutch participants did not understand what this question meant exactly (even after explaining multiple times) so they filled in 'neutral' which is converted into a 4 in the total scale of options. However in France, participants who had trouble understanding this question were informed to leave the question blank, which was then converted into a 0. For more insight into the exact results, please see Annex C.

2.5.2 PSSUQ Ratings

The PSSUQ measures multiple facets of the overall system by means of 18 questions. As there are four of these facets, the PSSUQ results will be described in the following subsections. There were seven options for answering ranging from "I totally agree" to "I do not agree at all", which were converted to numerical values. A lower value represents a higher level of satisfaction and vice versa. For an overview of the exact results, see Annex D.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 16 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

2.5.2.1 Overall Rating

One user from France left a number of questions blank. These questions were left out the calculations made. The overall rating was very positive, with an average rating of 2.04 (SD = 0.52). As this is below 3.5, users were quite satisfied about the system as a whole.

2.5.2.2 System Quality

The average value for the items 1 through 6, measuring system quality, was 1.52 (SD = 0.09). Average values for all questions were below 2. Users thought it was an easy system to learn and easy to work with, fast to work with and they felt comfortable working with it. Looking at the individual users, there was one user who appeared to have a bit more trouble with the system than the others, but her ratings were still in the positive range.

2.5.2.3 Information Quality

One participant left all question measuring information quality blank. Without her ratings, values indicate a mean of 1.98 (SD = 0.3). Users were still satisfied with the information quality, but less than with the overall system or the system quality. Looking at the individual user scores shows that one user in particular was not satisfied with the information quality of the system, giving an average rating of 6.5 on item 7 – 12. Furthermore, users were most troubled that the system did not provide error messages to fix a problem (M = 2.83, SD = 4.15) by clarity of information (M = 2.75, SD = 1.86), the help offered by the system (M = 2.75, SD = 1.82).

2.5.2.4 Internal Quality

Internal quality was rated positive, with an average value of 1.98 (SD = 0.3). People were especially positive about using the vocal interaction (M = 1.75, SD = 0.62). Participants had some doubts about the volume of the audio (M = 2.58, SD = 1.78), which also became clear when it was asked during the session whether the volume was sufficient enough.

2.5.2.5 Errors

Errors were captured by comparing the utterances of the participants (which were recorded) to both the predicted possible utterances and the possibilities for a good outcome. Errors were protocolled during the session by the test leader as well as after the session by studying the video documentation of the test session. All users were able to complete all tasks. For each individual task, errors will be discussed below.

2.5.2.5.1 Activate ALFRED

Most users didn't face any problems activating ALFRED. However two users had difficulties in unlocking the default screensaver on the smartphone. It was interesting to see that some participants wanted to activate the system by simply pushing the button while other participants (with smartphone experience) tried to swipe over the button for activation. One user tried to activate ALFRED by pushing the ALFRED symbol for a prolonged time continuously; prolonged pushing of the ALFRED symbol did not lead to an activation of ALFRED, but the participant then instead could have deleted the ALFRED symbol from the screen. Two participants did not know what the blue microphone is supposed symbolize. Furthermore, users differ in whether they deactivate ALFRED after a task or not. Some just forget or don't mind, while other users are really careful to

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 17 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

deactivate the system after each task. Three participants were confused with the red colour of the button (“does this mean it’s activated?”). Two participants did not know that the microphone button needs to be touched in order for ALFRED to listen.

2.5.2.5.2 Set an Alarm

The most occurring error (four out of nine participants) was that participants had a different choice of words. Instead of the word “alarm” or a sentence in which this word was used (e.g. “I would like to set the alarm”/“please set the alarm”), there were a variety of wording choices, ranging from “Please wake me up at...” to “I would like a wake-up call”. Another error that occurred often was not mentioning a date. Three users mentioned that the alarm had to be set on Tuesday, but without saying it had to be next Tuesday, or calling a specific date. One user even forgot to mention “Tuesday” and a date altogether. Other difficulties occurred because two users changed the order of the sentence completely (e.g. “next week Tuesday at 8:00 I would like to set an alarm please”). It is also mentionable (not an error in general) that three of the participants were really polite, saying that “they would really like to set an alarm” or asking ALFRED “could you please set the alarm?”.

2.5.2.5.3 Send a Message

The most common mistake, was that when users were asked to dictate the message, they dictated it as if they were talking to ALFRED, thus not realizing that this would be the message for the person in question (e.g. “please tell Steve...”, “Can you ask Steve..” or “I have a neighbour named Steve..”). This error occurred only among the Dutch users (all six of them) and German users (two users) and it might be caused by a language or translation difference in the prototype. Another problem two users encountered was that they had trouble hearing and understanding the sentence “please dictate your message”. As a result, one of these users turned off the system in confusion. Another participant turned off the system because she thought the message was sent already, while ALFRED only asked whether the message had to be sent. Last, one user just carried on with the next task without telling ALFRED that she wanted to send a message in the first place. Last, there was one user who did not use the word “message” but asked whether ALFRED “could order two bottles of milk for Steve”.

2.5.2.5.4 Call Someone

From all tasks, this was the task in which the least errors occurred. Two participants named the wrong person (someone from their personal life). Three participants were really polite again, asking ALFRED “could you call..?” or telling the system that they “would really like to call Lisa”.

2.5.2.5.5 Put a Reminder in the Agenda

One error that six participants made was that when they were asked to say a little about what the appointment was about, they wrote it down like they would tell it to a friend instead of how they would put it in a real agenda. They’d say for example “well...that I am going to meet my friend Steve for a drink”. One participant gave an exceptionally long explanation. Another error was that five participants forgot to mention the date for the reminder. One even forgot to mention a day altogether. Two other participants had a wrong intro (e.g. “hello ALFRED, could you remind my friend Steve..?” or something likewise).

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 18 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

2.5.2.5.6 Respond to an Automated Message from ALFRED

Something to take into account (not wrong per se) is that a lot of users tend to not just answer a question with a simple “yes” or “no”. When ALFRED asked whether the user wanted to hear more about the proposed event, four users replied with something other than “yes” or “no”, for example “please” or “tell me more”. The same was the case when ALFRED asked whether the participant would like to join the event: again, four users said something different (e.g. “I will come” or “that would be nice”). This is surely something to take into account for a later version of the prototype for the (real) system. One user did not immediately know how to respond, and wanted to push the button to hear more about the event. Another user had trouble hearing what was being said, and asked for a repetition of what ALFRED had said about the event.

2.5.3 Feedback from Participants

2.5.3.1 Positive Comments

Users were in general satisfied with the ease of use (mentioned by nearly all participants), as the idea of using a new technological device was rather scary to a lot of the participants. Also, they liked the simple design, it was mentioned by one user with little ICT knowledge that it was strange to operate a phone by voice at first, but after a short period of time it felt really natural and the user prefers voice interaction over typing or touching. The red waves around the microphone were perceived as a good indicator that ALFRED is actually listening. One participant said she really liked that ALFRED would propose nearby events, as this would combat social exclusion. Another participant noted that the application was really usable for people who had trouble using their hands, or searching for small buttons. A last participant said that she really liked the clear voice. However, as will be described below, opinions about the voice differed slightly in the group of participants.

2.5.3.2 Negative Comments

Difficulties in activating ALFRED were found by all users of the German pilot, additional information on the screen is necessary to give basic instruction on e.g. how to unlock the phone or how to activate and deactivate ALFRED. Two participants noted that it was weird that ALFRED had a female voice while the name “ALFRED” is rather masculine. One of these participants would rather like a male voice (“which is in general a bit lower, so I’ll be capable to hear it a bit better”). Two other participants thought that the voice sounded rather synthetic, which was not a good thing according to them. Yet two other participants had trouble understanding what ALFRED said, either because the volume was too low or because the voice was talking at a too fast pace for them. One participant mentioned that she missed a notification on the screen itself when activating ALFRED (the red waves around the microphone weren’t clear enough for this participant).

2.5.3.3 Recommendations

The recommendations section provides recommended changes and justifications driven by the participant success rate, behaviours, and comments. Each recommendation includes a severity rating. The following recommendations will improve the overall ease of use and address the areas where participants experienced problems or found the

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 19 / 62
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

interface/information architecture unclear. Recommendations are given by means of two ways: First, there are the recommendations leading from the errors that occurred per task. Second, participants gave some direct recommendations as well, right after the evaluation.

2.5.3.3.1 Task 1: Activate ALFRED

For the first task, users had to activate ALFRED. This could be done by either tapping the big red button in the middle of the screen or swiping over it. Table 3 below shows the errors participants encountered while trying to perform this task, proposed change to prevent this from happening, and the severity of this specific error.

Table 3: Feedback Summary for Task 1

Proposed change	Justification	Severity
Include an on-screen notification for when ALFRED is activated	Multiple users were confused to whether they had activated the application yet or not. Apparently, the red waves around the button are not a clear enough mark of activation for all users	Moderate
Make it possible to activate the application either by a push or a SWEEP	Inexperienced users will press the button while users with more Smartphone experience will try to activate the application by swiping	Low
Include a written instruction on the screen that explains how to deactivate the screensaver	Users with little ICT experience did not know how to unlock the phone and what the "lock symbol" did mean.	High
Long pushing of the ALFRED symbol also has to activate ALFRED	The user took rather long (60 sec) to activate ALFRED	High
Deactivate the "delete" function of ALFRED on the screen	Users might delete the ALFRED symbol	High
Include a written instruction on the screen close to the microphone symbol that explains either HTT or PTA	Users thought ALFRED was already listening after they clicked the ALFRED icon on the screen	Moderate

2.5.3.3.2 Task 2: Set an alarm

For the second task, participants had to set an alarm for the next day at 8:30 AM. Participants would just have to tell ALFRED to set an alarm and respond to further questions by ALFRED, if there were any. Errors encountered severity of these errors and proposed changes can be found in Table 4 below.

Table 4: Feedback Summary for Task 2

Proposed change	Justification	Severity
Make sure that ALFRED either knows that multiple words have the same meaning or create a stepwise process for activation in which users can only use certain words	A multitude of users use different words than "alarm".	High

for activation		
ALFRED must always ask a date if this is not asked or take the next day if only a certain day is mentioned and if the alarm should be am or pm (e.g. users says "Friday" without a specific "next" or date, then ALFRED will put the alarm next Friday)	Some users did not mention a day or date at all; impossible to set a proper alarm	High
Show the alarm as an entry in the calendar	The users want to see what other events they have planned on that day so they can manage their agenda in a more structured way	Low

2.5.3.3.3 Task 3: Send a message

Participants were asked to send a message to their neighbour Carl, in which they had to ask him to bring a bottle of milk from the grocery store where he works. Just as in the previous task, participants could do this by telling ALFRED they wanted to send a message. Then they had to respond to ALFRED if further questions were asked. Table 5 below displays the errors encountered severity of these errors and proposed changes.

Table 5: Feedback Summary for Task 3

Proposed change	Justification	Severity
Create a safety option for ALFRED: When a user tries to turn off the system while being in the middle of sending a message a question like "the message has not been sent yet. Are you sure you want ALFRED to shut down" will be asked	Some users turned ALFRED off in confusion without sending the message	High
Replace certain wordings that were hard to understand (dictate) with everyday words	Some of the users did not understand what ALFRED told them to do and were confused by this	Moderate to high
Make the process more stepwise (so the user will understand that the spoken message will go directly to the receiver), or let the user hear or see the message before sending ("this is what Steve will hear...").	Participants, the Dutch in particular, talk to ALFRED as a mediator for messaging, not as if they are directly talking to the person they want to send the message to	Moderate to low (message will still be send, it'll just be weird for the receiver)

2.5.3.3.4 Task 4: Call someone

In the fourth task, participants had to call someone. To do so, they just had to tell ALFRED who they wanted to call. Below in Table 6 are the errors, severity and proposed changes for this task.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 21 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

Table 6: Feedback Summary for Task 4

Proposed change	Justification	Severity
ALFRED must recognize politeness. So, "I would like to.." or "could you please..." Best to use code words such as "call". If the word "call" or something likewise is mentioned, ALFRED can ask whether the user wants to call someone	A multitude of elderly participants were very polite when talking to ALFRED	Moderate

2.5.3.3.5 Task 5: Put a reminder in the agenda

Next, participants had to put a reminder in the agenda for meeting up with a friend, Steve, for a drink Tuesday next week at 14:00. Participants had to tell ALFRED that they had a meeting, with whom, when (date and time), and what the purpose of this meeting was. ALFRED would ask questions if participants did not tell all the information at once. Table 7 below shows the errors, severity and proposed changes for task 5.

Table 7: Feedback Summary for Task 5

Proposed change	Justification	Severity
ALFRED must always ask a date and time am or pm if this is not asked or take the next day if only a certain day is mentioned (e.g. users says "Friday" without a specific "next" or date, then ALFRED will put the reminder on next Friday)	Some users did not mention a day or date at all; impossible to create a reminder in the agenda	High
Make the process more stepwise (so the user will understand that the spoken message will be placed directly in the agenda), or use the word "title" instead of asking for a description	Participants, the Dutch in particular, talk to ALFRED as a mediator for messaging, not as if they are directly talking to the person they want to send the message to	Moderate to low (title will still be in the agenda, it will just be weird)
Display the reminder on a calendar on the screen along with the users agenda	Users said they cannot put in an reminder if they don't see what other things they have planned on that day	High

2.5.3.3.6 Task 6: reaction to an automated message from ALFRED

For the final task, participants had to wait until ALFRED would send out an automated message. In this message, ALFRED asked the participant whether he or she would like to go to an event nearby, where a friend would go to as well. Participants had to reply, either by telling ALFRED that they didn't want to go or that they would like to go to the event. ALFRED would respond with a question regarding subscribing the participant to this event.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 22 / 62
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

The errors encountered by users, severity of these errors and proposed changes for this task is listed below in Table 8.

Table 8: Feedback Summary for Task 6

Proposed change	Justification	Severity
Make it possible to turn automatic messages off, or on stand-by	A lot of participants (5 out of 6 Dutch participants) indicated that they do not always want to receive a message from ALFRED (e.g. "not when I am in the bus or with my friends. I would find that very annoying")	High (if this annoys the user, they won't take ALFRD with them anywhere)
Make sure ALFRED understand multiple wordings instead of just "yes" and "no", or indicate to the user that the answer to a question should be "yes" or "no"	A multitude of participants answered with something else than a mere "yes" or "no" when they were asked whether they would go to a proposed event	High – moderate
Create a sound for messages by ALFRED	User likes to know by sound alone that the message comes from ALFRED	Low
Show the suggested event on a calendar on the screen along with the users agenda	Users said that they cannot confirm an event if they do not know what other things they have planned on that day,	High

2.5.3.3.7 Other recommendations

Participants also encountered some incidents that were not specifically related to a task. Some participants also proposed some improvements or preferences by themselves. These proposed changes and their justification can be found in Table 9 below.

Table 9: General Feedback Summary

Proposed change	Justification	Severity
Set the default volume higher	Multiple participants did not always hear what ALFRED was saying, even at the loudest volume	Moderate – high (if users can't hear what is being said, it will be very difficult to communicate with ALFRED)
Create a stand-by possibility for ALFRED (maybe with a code word, just like the Dutch alarm number 112 always works, even if your phone is locked)	At least two users indicated that they would like this for safety reasons. So that when they would fall and be unable to reach for their mobile device, they could still tell ALFRED to call someone even without having to push the button for activation	Moderate – high (users deemed this very important)
Implement a possibility so ALFRED can speak slower if the user wishes so	Some participants had trouble understanding ALFRED because of the fast speaking pace.	Moderate
ALFRED must know the comment "repeat"	Users asked ALFRED to repeat what was just said because they did not understand it the first time	Moderate
Implement a possibility to	User preferences and an inconsistency with a	Low

change the voice to a male voice	male name/female voice	
Implement a “you’re welcome”	Users tend to say “thank you” very much, and “you’re welcome” sounds nicer (and more logical) than “goodbye	Low
Know the word “no”	If the uses is misunderstood or does not want to do a suggested option or task, ALFRED must recognized that	Moderate
ALFRED must know the comment “back”	If the user has given a command the he wants to change, saying the word “back” must allow him to go back to the previous utterance	Moderate
ALFRED should give a warning before he switches into “Standby” mode	Users got confused and did not know if ALFRED was still listing after the phone switched into “Standby”. They stoped talking to ALFRED as they thought ALFRED has switched off.	Moderate
Apply a ribbon to the phone, so ALFRED can hang from the neck of the user	One user wanted to use the ALFRED hands free while doing her work at home	Low
Give a reminder that PTA is still activated after 5 min	If the user forgets to switch of PTA a reminder that ALFRED is still listing was desired for privacy reasons	Low
Increase speaker output of the phone	The volume was too low for one participant, even at maximal settings	Moderate

2.6 Conclusion

The evaluation test gave a good overview of people’s first impression of a system that can be controlled by voice interaction. Most users were very enthusiastic about it, as they immediately saw the benefits of a system that has no annoying small buttons or difficult menus. This is also visible from the results in both the ASQ and the PSSUQ which indicated a high level of satisfaction among participants.

Users were in general really satisfied with the system and the idea of voice interaction. Also, most of the users experienced little to no real problems while working with ALFRED. However, some important findings for improvement were done.

- It was not always clear for participants that ALFRED was already activated
- Participants were mostly not too fond of the ALFRED voice for multiple reasons
- Participants use a multitude of different words which mean the same thing
- Participants sometimes forgot to mention a specific date or day for appointments
- Participants talked to ALFRED as if it was a person (e.g. when sending messages)
- Participants want to be able to turn off the ALFRED messages in some instances

It is difficult to imagine how a real system would respond to this, as a person being the Wizard in this evaluation would approve of utterances a system probably would not comply with. Human speech is highly complex, something that was already visible. Implementing

the recommendations into a real (non-WOZ) system and continuing to work with users (i.e., real lay persons) will ensure a continued use of ALFRED.

Finally, as the D8.1.1 states in the chapter 2.2, the iterative evaluation is a continuous process that will be carried out throughout the project. The next step is the first prototype evaluations that were initially planned to be carried out in the Netherlands, Germany and France between the M18 and M23, but this should be adapted according to the first prototype deliveries (of the different ALFRED components) from the technical project partners. The iterative testing methodology that was put in place for the first evaluations will be adapted and updated according to the components that are tested. This work will be reported in the forthcoming deliverables of the WP8.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 25 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

3 Pilot methodology

3.1 Methodology

The aim of the ALFRED pilots is to demonstrate the real-world applicability of ALFRED and to evaluate its impact [DoW]. Each ALFRED pilot has specific characteristics and objectives that work towards this common aim within a holistic evaluation framework.

The holistic evaluation framework is based on different perspectives. ALFRED is focused on ICT for independent living with very broad objectives. Current models for trial research on ICT for independent living focus either on usability studies, clinical validations, technology use in real life or impact research. However a combined model that is suitable for ALFRED is not available.

Therefore the project will adapt and combine existing methodologies, tools and KPI's within a holistic framework of perspectives. The overall framework is depicted in Figure 2 and focuses on four perspectives. The main aim of these perspectives is to provide a common validated framework to guide the pilots towards a common goal as will be described in the following section. This section presents each perspective with a short indication on evaluation methods and KPI's that will be further extended in the specific pilot sections (section 4).

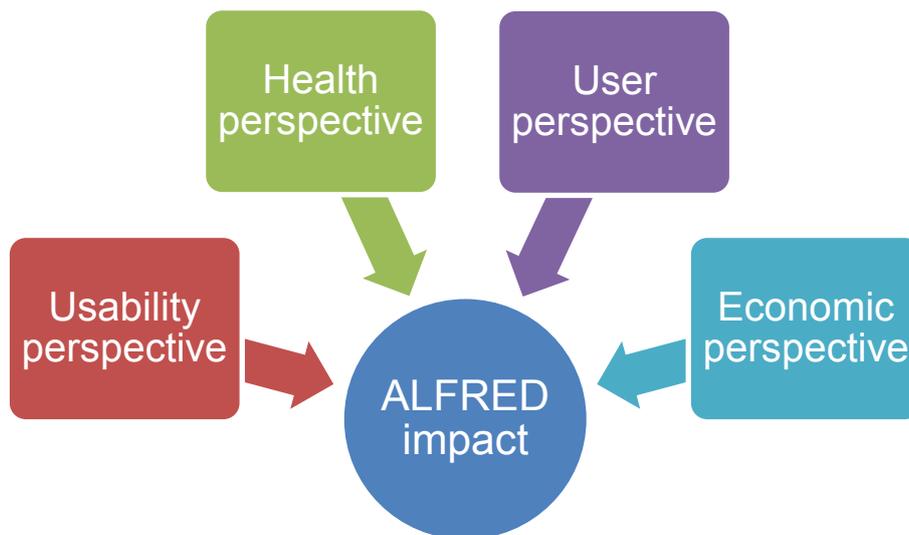


Figure 2: Overall Framework of ALFRED Impact

3.1.1 Outline of the Pilot Perspectives

The iterative testing process is focused on the usability perspective of the different components in the ALFRED-system. In the pilot phase, three additional other evaluation perspectives are included in order to assess the ALFRED system as a whole.

The following figure depicts how the different perspectives interrelate and in which pilot countries they will be analyzed. The figure reflects how the perspectives are all interrelated and that they are tackled in the different pilots. The pilots have been organized according to their capacities and profile in order to work on the different perspectives and reach a

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 26 / 62
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

common goal. The following sections give the overall planning of the pilots and the pilot specific sections (section4) will further define the evaluation methods and related KPI's.

The following figure demonstrates where the different perspectives will be performed. Please note that Pilot 1 is in the Netherlands, Pilot 2 in Germany and Pilot 3 in France.

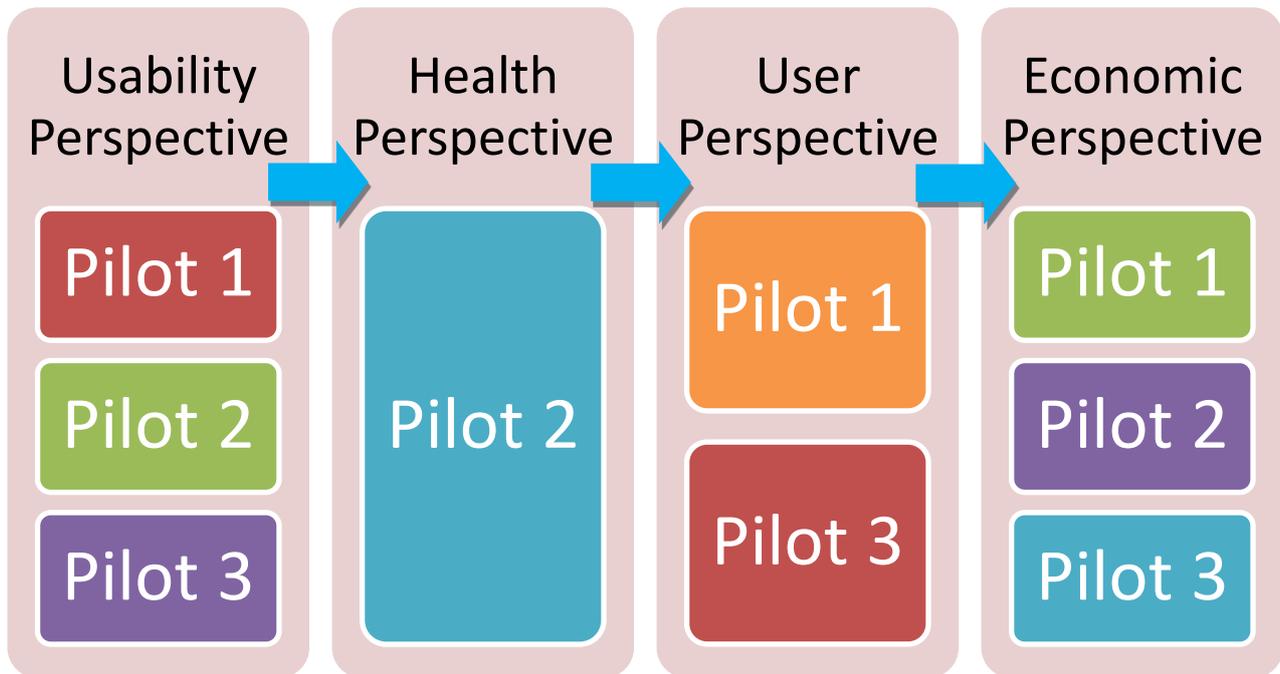


Figure 3: Infographic – Overview of the Pilot Plan

In the following subchapters, the four different evaluation perspectives are detailed and the first methods of evaluation foreseen. However, it is important to understand these methods will be specified as the system prototype development moves forward and the iterative testing phase gives more results. In consequence, these methods of evaluation for each pilot perspective are updated in the forthcoming WP8 deliverables (notably D8.2.1). Hand in hand with the methods of the evaluation; the KPIs described in the following sub-sections will be updated when necessary.

3.2 Usability Perspective

When developing an ICT solution for older people, the ease of use of a solution is of crucial importance. Therefore the first perspective focuses on usability of ALFRED, involving firstly older end users in the usability evaluations as a continuance of the iterative evaluation sessions (see 8.1), and secondly primary and secondary users in the pilot phase. With this research, that has already started, the ease of use for the user will be defined by evaluating its usability. Usability is the extent to which a product can be used to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use [Nie94]. This usability research will be performed in all ALFRED pillars. The different pilots divide the usability research among each other to focus on the speciality of each partner. The Dutch pilot focuses on pillar I and II, the German pilot on pillar III and IV and the French pilot on pillar I, II and IV), although there may be some overlap when this is convenient for research purposes.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 27 / 62
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

3.2.1 Envisaged Methods for Evaluation

As explained in D8.1.1 and as presented in this deliverable, the usability perspective will be measured by applying the same method used before. A combination of observation and think aloud protocols will be implemented with the Post Study System Usability Questionnaire (PSSUQ) and After Scenario Questionnaire (ASQ) [Lew91], [Lew95]. The PSSUQ is an 18-item questionnaire that measures the perceived users' satisfaction with a product or system. Obtaining an overall satisfaction score is done by averaging the four sub-scales of System Quality, Information Quality and Interface Quality. The PSSUQ is highly reliable (.94) and available free of charge. The After Scenario Questionnaire (ASQ) [Lew95] is a three item usability questionnaire that will be given to subjects directly after they have completed a task. These questionnaires will remain identical for the usability testing in ALFRED in order to obtain a linear effect throughout the development of the different prototypes.

3.2.2 Usability KPI's

Table 10 describes the Key Performance indicators that are foreseen to be used.

Table 10: KPI for the Usability Evaluation

Concept	KPI	Strategic Goal
Ease of use of the final ALFRED system	Average amount of errors (unable to complete the task) during the last usability cycle with the entire ALFRED system.	Less than 3
Adaptation to the needs of older end users and uptake of end user input.	Rise of PSSUQ and ASQ satisfaction levels towards the final usability cycle	Final increase
User satisfaction with the ALFRED system	Percentage of positive feedback in each of the validation groups after the last usability cycle.	60%

3.2.3 Output and Upcoming Deliverables

The deliverable D8.2.1, which is due in M28 two months after the end of the iterative testing process, will reflect the results of the usability evaluations that were performed from M13-M26. It will furthermore provide the final piloting definitions. Based on these evaluations on the ALFRED target group, a guideline or checklist will be created which will help developers to create apps for older adults, using speech interaction. This checklist will be part of a separate deliverable (D8.2.2) because of its market value.

3.3 Health perspective

Aside the usability and socio economic perspectives, this evaluative research focuses on the impact on health and physical wellness and prevention. Physical wellness and prevention involve aspects of life that are necessary to keep older adults in a good condition and can limit the effects of age related physical conditioning, like muscle wasting, reduced balance, and decrease in cardio-vascular endurance, osteoporosis and cognitive function. While physical activity has a positive effect on all of these aspects, not all of these will be individually assessed in ALFRED, since some diagnostic methods will involve

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 28 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

radiation burden to the users and are therefore out of scope. In this pilot, ALFRED aims to empower users to be able to monitor their own vital signs and understand their body's warning signs and give users the opportunity to perform health related monitoring and training tasks independently at home or in a healthcare setting. The research makes sure that older end users can use the ALFRED health related apps safely and independently at home and develops a checklist that health related apps must abide to in order to guarantee safety. A detailed description of the role of CHA within the pilot can be found in D6.0 "Clarification of WP's scope" in section 2.3.

3.3.1 Envisaged Methods of Measurements

Monitoring of the vital parameters temperature, pulse and respiratory rate are common in the physical wellness and prevention sector. Measuring these values help to guarantee compliance with safe and effective exercise guidelines and leads to optimal training effects. These effects will be measured by a variety of non-invasive and non-radiation based validated medical assessments and questionnaires. Measurements will include the increased amount of physical activity that older adults are performing during the pilot while using the ALFRED systems well as the effects that are related to preventive and physical wellness related goals. Examples for measuring the effects of ALFRED on users in pilot 2 are changes in mental wellbeing, sleep quality, H₂O consumption, physical activity and the ability to perform activities of daily life. A more detailed insight on the effects of the physical wellness can be gained by the medical assessment like Berg Balance scale (BBS). The BBS is made up of 14 functional task items out of common day life that progress from sitting to standing, standing to sitting, picking up an object of the ground, turning 360 degrees until standing on one leg. The BBS was designed to assess the balance ability of the elderly. The 5 times sit to stand (FTSST) is one method to evaluate how well older adults can rise from a chair, a task performed multiple times daily by active people and is a basic measurement for muscle strength and the ability to perform activities of daily life. Grip strength is an assessment tool in evaluation of the elderly and also a valid and reliable tool for evaluating upper extremity strength impairment. The 6 minute walking test evaluates the distance a subject can quickly walk on a hard, flat surface over the said period of time and therefore can provide information on the cardiac stamina of the user. The Falls efficacy scale measures the fall risk of older adults. Overall activity can be measured with the aid of pedometers give insight on physical activity behaviours of patients. One classification, the graduated step index, including only quantity of steps per day creates four categories. Category one would be less than 5000 steps per day and is labelled a sedentary lifestyle, in between 5000 and 7500 steps would be considered low active, up until 10000 is considered active and more than 12 5000 steps would be considered highly active. In older adults a recent review [Tud et al 11] found healthy subject to take 2000-9000 steps per day and a guiding value of 3000 steps was suggested. A comprehensive list with detailed descriptions on the possible measurements that are envisaged in Pilot 2 can be found in the D6.1 "User Health Profile and Carer Relationship Definition". The final measurement tools for the evaluation of this pilot will be described in D8.2.1 and will be dependent on the exact technical specification of the pilot hard- and software.

This perspective will be analysed within the German pilot in the scope of pillar III and IV, due to its expertise in the health and clinical research.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 29 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

3.3.2 Health KPIs

Possible KPIs to be measured in Pilot 2 are increased amount of physical activity (SF-36 subscale 3-KÖFU), increased mental wellbeing (Warwick-Edinburgh Mental Well-being Scale or PHQ-9) by 10%, increase in sleep quality by 10% (Pittsburgh Sleep Quality Index (PSQI)), and increased amount of H₂O consumption of 10%.

Table 11: KPI for the Health Care Evaluation

Concept	KPI	Strategic goal
ALFRED II and III effects on physical activity for older people	Increased amount of physical activity by 10%	Average increase of 10%, during the German pilot
ALFRED II and III effects on physical activity for older people	Increased mental wellbeing by 10%	Average increase of 10%, during the German pilot
ALFRED II and III effects on physical activity for older people	Increase in sleep quality by 10%	Average increase of 10% during the German pilot
ALFRED II and III effects on physical activity for older people	Increased amount of H ₂ O consumption of 10%,	Average increase of 10% during the German pilot

3.3.3 Output and upcoming deliverables

The deliverable D8.2.2 will contain the results of iterative evaluations. D8.3.2.1 will give the final specifications for pilot 2, which will depend on the technical progress and the exact specifications of the available serious games. The final results of pilot 2 will be provided in the deliverable D8.3.2 in M36.

3.4 Economic perspective

In addition to the usability evaluations, the pilots aim to reflect the economic value of ALFRED after the end of the project. The economic impact of Alfred can be measured through its potential to reduce healthcare burden and costs while maintaining or even improving the quality of care. Prevention and wellness can be a support for ALFRED users to improve their quality of life through monitoring health indicators, social support, health education, improving communication with caregivers, awareness and behavioral changes. This economic perspective will be approached in two ways. First of all, as ALFRED aims to support effective and personalized care within pillar III. It supports carers and healthcare staff to achieve more effective care. Therefore the impact will be measured on a reduction of costs for healthcare providers. In particular the costs related to reduction of staff through monitoring on distance. This study will be performed in a clinical surrounding to evaluate the actual cost benefit for formal caregivers in the German pilot (a hospital) and serve as input for the business case.

Secondly the economic perspective from the end user point of view will be tackled in the Pilot 1, the Netherlands and Pilot 3, France. These pilots will study up to what level older end users can and are willing to pay for the added value offered by ALFRED. This will also give important insights for the business case.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 30 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

3.4.1 Envisaged Methods of Evaluation

Evaluation of the time consumed per visit of a professional healthcare provider. This will be assessed by comparing the standard durations of time which are needed by a healthcare provider when dealing with healthcare related tasks in the area of prevention. These tasks can include time spend on supervision of the user, preparation of exercise material, documentation and total number of visits. A proportion of the tested users will be equipped with sensors to measure heart rate, body temperature and breathing frequency. These sensors might be confidence inspiring and it will be assessed with means of a questionnaire whether these sensors lead to reduction of visits to healthcare providers.

Secondly an analysis will be made on the different options for selling ALFRED together with end users after they have used the ALFRED solutions independently at home. In order to obtain this input, a qualitative study will be performed through a structured interview with the older target group. The structured interview will be based on the Business Model Canvas of Osterwalder [Ost08]. A structured interview will make it possible to explain and discuss different ALFRED business model aspects with the end user and discuss different options. A first version of the structured interview will be available in D8.2.1 and this will be further refined based on results of WP9.

The German pilot will research the socio economic implication of ALFRED healthcare apps on cost benefits in a hospital surrounding. In the Dutch and French pilot a socio-economic research will be made focusing directly on the older end-user and what they are willing to invest.

3.4.2 KPIs

Table 12 lists the identified KPIs for economic evaluation perspective. It start first with an overview of KPI's within the healthcare structure and secondly KPI's end user economic evaluation.

Table 12: KPI for the Economic Evaluation

Concept	KPI	Strategic Goal
ALFRED impact on healthcare costs	Reduction of the number of visits to healthcare providers for the purpose of sensor checks.	10% during the German pilot
ALFRED impact on healthcare costs	Reduction of the time that will be required for each visit to a healthcare provider	10% during the German pilot
ALFRED impact on healthcare costs	Reduction in self referred visits to a healthcare professional because of easy access to users own health parameters	10% during the German pilot
ALFRED impact on end user willingness to pay.	Percentage of TPs interested in buying the ALFRED system (either as a package or as an app on their smartphone).	50%

ALFRED impact on end user willingness to pay.	Average amount TP's indicate that they are willing to pay for an ALFRED package (Smartphone and Apps, excluding sensors).	More than 200 ¹ €
ALFRED impact on end user willingness to pay.	Average amount TP's indicate that they are willing to pay a month for leasing an ALFRED package (Smartphone, apps, sensors, 3G connection)	More than 30€ a month

3.4.3 Output and Upcoming Deliverables

The output of this research will be important input for D9.1 on Exploitation. Additionally the results will be reported in the final pilot deliverables.

3.5 End user Perspective

This perspective aims to reflect the added value of ALFRED for primary and secondary end users. It will define the acceptance of the ALFRED system and the added value it provides to the lives of target users. This will be defined by implementing the ALFRED system at the homes of older end users, where they can use it during a longer period of time. The ALFRED system will be used both by primary (older people) and secondary (informal caregivers) end users. This will be done after pilot 2, Germany (where primary and secondary users² use ALFRED in a controlled environment), has confirmed that the ALFRED system can be used safely in the individual home environment.

The perspective will indicate whether users will keep on using ALFRED and are willing to pay for it (see the economic perspective). Additionally the perspective will define whether ALFRED can be related to an improvement in quality of life. Although the direct causal link between ALFRED and an improvement in quality of life will be difficult to prove within the limited time frames of the project, it is included in the end user perspective to at least provide for some first insights and provide for possibilities to cross reference between different results.

3.5.1 Envisaged Methods of Measurement

In order to define the acceptance and added value of ALFRED for the target users, the Technology Acceptance Model will be used [Davis, 1989]. The Technology Acceptance Model (TAM), will help us to predict the acceptance and use of ALFRED by older end users. The original TAM involves perceived ease of use, perceived usefulness, behavioural intention and behaviour as determinant factors for the acceptance of technology (Figure 4).

¹Price is derived from similar solutions on the market (see <http://www.paradigit.nl/smart-life/golivephones>) and is only used to concretize the indicator.

² For more information about the ALFRED target user segments, see the updated version of D2.3 User Stories and Requirement analysis (new section 2.2.4)

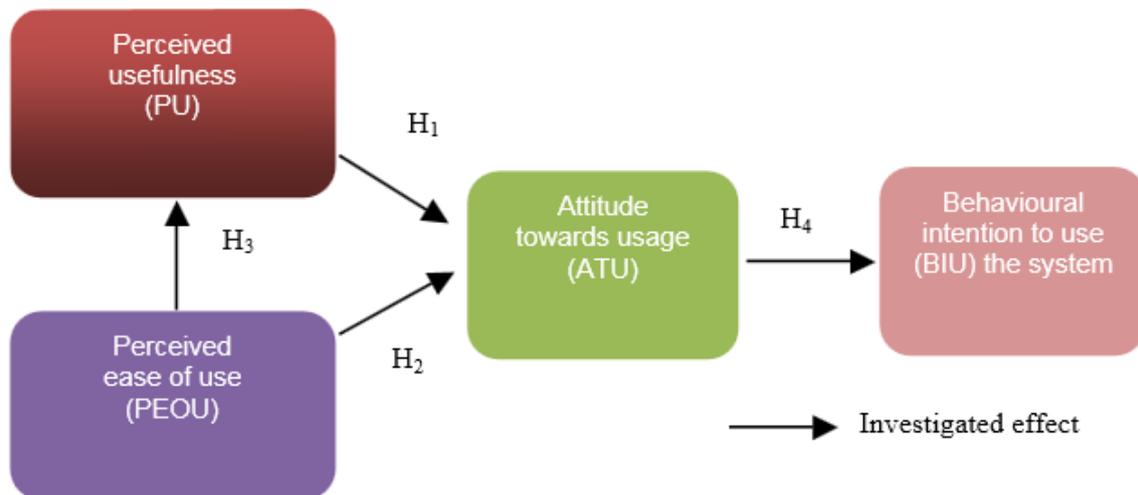


Figure 4: Technology Acceptance Model [Davis, 1989]

The perceived ease of use and usefulness in this model will be measured for the ALFRED system with the help of the previous usability study using the ASQ and PSSUQ (section 3.2.1). The behavioural intention will be measured with a standard questionnaire on the intention of use. In order to compare the intended behaviour with the actual behaviour of the participations the use will be monitored with log files during the pilot. This will be explained to the test person (TP) in the informed consent. The log files will be complemented with a diary that the TP can use. Experience shows that older TP find it difficult to keep a diary and that this is unreliable, therefore the log files are complementary to this. The comparison between the intended use and actual use will make it possible to draw conclusion on the added value of ALFRED for older people. This results can then be compared to a possible improvement in quality of life. Although it will be difficult to establish a causal link with improvement of quality of life, due to the small sample and short timeframes, the results can be used to cross reference between each other.

In order to measure an improvement on quality of life, first of all a zero measurement will be made on quality of life at the beginning of the pilot. The Older People Quality of Life Survey [OPQOL-35] will help to define the zero measurement before the use of the ALFRED system. After a month of use the survey will run again to define the perceived impact on quality of life. This will help to define whether ALFRED supports the ability to perform daily activities and promotes autonomy of older persons.

The research method will be developed in a practical workbook that can be used to run the pilots at the homes of older people [Deliverable 8.2.1]. The perspective will be analyzed within the Dutch and the French pilot using all ALFRED pillars at the homes of older people independently.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 33 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

3.5.2 KPIs

Table 13 lists the KPIs that are relevant for the evaluation from the perspective of the users.

Table 13: KPI for the End User Evaluation

Concept	KPI	Strategic goal
Acceptance of ALFRED by older end users - User Driven Interaction Assistant	Average amount of messages sent with ALFRED per week to contacts	5
Acceptance of ALFRED by older end users - User Driven Interaction Assistant	Average amount of calls made with ALFRED to contacts per week	5
Acceptance of ALFRED by older end users - Social inclusion	Average amount of additional participation on social events (e.g. visited an event I would otherwise not have done), during the pilot	3
Acceptance of ALFRED by older end users – Effective and Personalized Care	Average amount of times an end user measures and checks one of his vitals at home during the pilot	6
Acceptance of ALFRED by older end users – Health monitoring	Average amount of times a secondary end user checks on the vitals of an older person	4
Acceptance of ALFRED by older end users – Physical and Cognitive Impairments Prevention	Average amount of times an older person play a game per week	3
Perceived ease of use and usefulness for older end users	Actual use larger than intended use (average)	>10%
Impact on quality of life and independent living	Increase in perceived improvement on Quality of Life (without guarantees that this is directly linked to ALFRED)	>10%

3.5.3 Output and upcoming Deliverables

The results of this research will be available in D8.2.3, D8.2.4 and in D8.4 on the pilot reports.

3.6 Overview Pilot Methodology

To clarify the previously discussed Pilot Methodology, this section gives a short overview of the evaluation methods in the scope of the perspective, the pilots and the pillars.

3.6.1 ALFRED Pillars and Pilots

Table 14 below gives a short overview of what methodology (questionnaires, scales, assessments, etc.) will be used for each perspective and in which pilot and pillar it will be used. This will also be further clarified in section 4 on the individual pilots.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 34 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

Table 14: Initial ALFRED Pilot Methodology Overview

Perspective	Methodology	Pilot	Pillar
Usability perspective	After Scenario Questionnaire [Lew95]	All pilots	All pillars
Usability perspective	Post Study System User Questionnaire [Lew02]	All pilots	All pillars
Usability perspective	Observation [HD06]	All pilots	All pillars
Usability perspective	Think aloud [HD06]	All pilots	All pillars
Economic perspective	Healthcare expenditure reduction	Germany	Pillar III and IV
Economic perspective	Structured user interview based on aspects of Business Model Canvas [Ost08]	France, the Netherlands	All pillars
Healthcare scientific perspective	Pittsburgh Sleep Quality Index	Germany	Pillar III and IV
Healthcare scientific perspective	Physical Activity Questionnaire (SF-36 Subscale 3 KÖFU)	Germany	Pillar III and IV
Healthcare scientific perspective	Warwick-Edinburgh Mental Well-being Scale	Germany	Pillar III and IV
Healthcare scientific perspective	H ₂ O consumption	Germany	Pillar III and IV
End user perspective	Older People Quality of Life [OPQL-35]	France and the Netherlands	All pillars
End user perspective	TAM methodology [Davis 1989]	France and the Netherlands	All pillars
End user perspective	Personal testing experience diary	France and the Netherlands	All pillars
End user perspective	Personal System activity log files	France and the Netherlands	All pillars
End user perspective	Pilot workbook	France and the Netherlands	All pillars

3.6.2 KPI Overview

The key performance indicators accompany the pilot evaluation methods and enable to assess the success of the evaluation phase and the impact the ALFRED system during the pilot phase. Table 15 lists the KPIs that have been foreseen so far in the pilot planning, this list can be updated in the forthcoming WP8 deliverables.

Table 15: Overview of the KPIs Used in the Pilots

Concept	KPI	Strategic Goal
Ease of use of the final ALFRED system	Average amount of errors (unable to complete the task) during the last usability cycle with the entire ALFRED system.	Less than 3
Adaptation to the needs of older end users and uptake of end user input.	Rise of PSSUQ and ASQ satisfaction levels towards the final usability cycle	Final increase
User satisfaction with the ALFRED system	Percentage of positive feedback in each of the validation groups after the last usability cycle.	60%
ALFRED II and III effects on physical activity for older people	Increased amount of physical activity by 10%	Average increase of 10%, during the German pilot
ALFRED II and III effects on physical activity for older people	Increased mental wellbeing by 10%	Average increase of 10%, during the German pilot
ALFRED II and III effects on physical activity for older people	Increase in sleep quality by 10%	Average increase of 10%
ALFRED II and III effects on physical activity for older people	Increased amount of H2O consumption of 10%,	Average increase of 10%, during the German pilot
ALFRED impact on healthcare costs	Reduction of the number of visits to healthcare providers for the purpose of sensor checks.	10% during the German pilot
ALFRED impact on healthcare costs	Reduction of the time that will be required for each visit to a healthcare provider	10% during the German pilot
ALFRED impact on healthcare costs	Reduction in self referred visits to a healthcare professional because of easy access to users own health parameters	10% during the German pilot
ALFRED impact on end user willingness to pay.	Percentage of TPs in France and the Netherlands interested in buying the ALFRED system (either as a package or as an app on their smartphone).	50%
ALFRED impact on end user willingness to pay.	Average amount TP's in France and the Netherlands indicate that they are willing to pay for an ALFRED package (Smartphone and Apps, excluding sensors).	More than 200€
ALFRED impact on end user willingness to pay.	Average amount TP's in France and the Netherlands indicate that they are willing to pay a month for leasing an ALFRED package (Smartphone, apps, sensors, 3G connection)	More than 30€ a month
Acceptance of ALFRED by older end users - User Driven Interaction Assistant	Average amount of messages sent with ALFRED per week to contacts	5 times/week in the French and Dutch pilot
Acceptance of ALFRED by older end users - User Driven Interaction	Average amount of calls made with ALFRED to contacts per week	5 times/week in the French and Dutch

Assistant		pilot
Acceptance of ALFRED by older end users - Social inclusion	Average amount of additional participation on social events (e.g. visited an event I would otherwise not have done), during the pilot	3 times in the French and Dutch pilot
Acceptance of ALFRED by older end users – Effective and Personalized Care	Average amount of times an end user measures and checks one of his vitals at home during the pilot	4 times in the French and Dutch pilot
Acceptance of ALFRED by older end users – Health monitoring	Average amount of times a secondary end user checks on the vitals of an older person	4 times in the French and Dutch pilot
Acceptance of ALFRED by older end users – Physical and Cognitive Impairments Prevention	Average amount of times an older person play a game per week	3 times/week in the French and Dutch pilot.
Perceived ease of use and usefulness for older end users	Actual use larger than intended use (average)	>10%
Impact on quality of life and independent living	Increase in perceived improvement on Quality of Life (without guarantees that this is directly linked to ALFRED)	>10%

3.7 Practical Implementation of the Pilot Methodology

In order to implement the perspectives and research activities, the following describes the short practical implementation of the Pilot Methodology.

3.7.1 Time Plan

The following table gives an overview of the time plan for the implementation of the pilots in the different pilot countries.

Table 16: Overview Piloting Time Plan

Month	Pilot Task	Details
M12-M26	Iterative Testing Evaluation	This testing helps to prepare the actual pilot phase (define the most relevant evaluation tools, get in touch with interested end-users etc.)
M23-M26	Final Usability Research with 20 end users on the usability perspective to validate the overall ALFRED system.	This testing will be a preliminary phase to running the ALFRED system at the homes of older people.
M26-M36	The pilot in Germany will evaluate with 40 end users the health perspective and socioeconomic perspective.	This testing will partly validate the health aspects before the pilot running at the homes of older people.
M29-M36	The pilot the Netherlands on all pillars with 10 end	

	users on end user and socio economic perspective	
M29-M36	The pilot the France on all pillars with 10 end users on end user and socio economic perspective	The testing will validate the ALFRED system as a whole in a day-to-day usage in home environment.

3.7.2 Recruitment

The following section describes the involved test participants in each country.

3.7.2.1 Profile Test Participants

Test Participants (TP) are older persons (60+) with different socio-economic backgrounds and with different levels of computer literacy and experience of mobile applications.

The Test Participants will be selected from the target group of ALFRED (see D2.3) through the networks of the end user partners according to their recruiting processes. The TP will give consent to participate in the pilots. Table 17 details the different target group segments that will be participate in the specific pilots.

Table 17: Overview of Involved Target Groups in the Different Pilots

Pilot	Country	Test Persons from the following target user groups ³
Pilot 1 – Individual Usability	Netherlands	PTG1, PTG2, PTG3 STG1
Pilot 2 – Hospital Environment	Germany	PTG1, PTG2, PTG3, PTG4 STG1, STG2
Pilot 3 – Day-to-day Usage	France	PTG1, PTG2, PTG3 STG1

Before the start of the pilot, there will be an information session where the TP will be informed on all the implications of participation in the pilot. The user will fill in the pre-questionnaire.

3.7.2.2 Amount of Test Participants

The following table describes the estimated number of participants that will be involved during the testing process in ALFRED. While the total amount of will be dependent on the available number of sensors and will be described in more detail in D8.2.2. It is currently planned to involve persons from the Secondary Target Group (STG) in the iterative testing phase, the decision the include STG users will be made dependent on the technical maturity of ALFRED at the early stages of the project.

³ See the updated version of D2.3 User Stories and Requirement analysis (new section 2.2.4) for the specifications of the target user segments

Table 18: Amount of Test Participants in the Pilot Countries

Pilot	Involved TPs from the Primary Target Groups (PTG)	Involved TPs from the Secondary Target Groups (STG)	Total number of the involved end-users
Iterative Testing Phase (three countries)	Netherlands: 5-6 at each cycle Germany: 3 at each cycle France: 3 at each cycle	-	At least 11 per cycle
Pilot 1 (Netherlands)	20	2	22
Pilot 2 (Germany)	37	3	40
Pilot 3 (France)	8	2 informal caregivers	10

3.7.2.3 Drop-out Strategy

This section discusses some techniques that trial design and management teams can use to reduce the frequency of dropouts. First, designers and managers can limit participants' burden and inconvenience in the data collection stage. This can be done in several ways. Within the ALFRED pilots, emphasis is put on objectives like minimizing the number of visits and assessments, collecting only the information that is needed at each visit, using user-friendly report forms and information sheets, using direct data capture that does not require a personal visit whenever feasible. Examples of information not needed at each visit include aspects of the participant's health history and contact information that were provided at earlier visits and information available from health records. The overall aim is to balance the competing goals of reducing response burden and collecting sufficient information to fully support the analytic goals and to guide the next steps in treatment. (Regarding use of direct data capture to minimize the response burden, it would also be useful to attempt to collect whatever information is available from administrative records.)

Initial introduction in for the Pilot phase designed for the Test Participants can also emphasize the importance of the informed consent process as a mechanism for ensuring that participants understand the commitment they are making, including their intent to complete the pilot. Training of the test administrators and research staff should also emphasize how to work with participants to minimize the extent of missing data. Finally, the end user partners need to know and explain to participants how to handle ALFRED within the scope of the pilot in a sufficient way and give the end users the possibility to answer all relevant questions before beginning the actual pilot.

The persons who drop out of the study will be short interviewed briefly to understand the reasons for the drop out and to improve the pilot experience for the other participants.

3.7.3 Helpdesk

With the use of a help desk the end user organizations will provide the end user with information and support related to ALFRED and the functions related to the individual pilots. The purpose of a help desk is to troubleshoot problems and provide guidance about products such as mobile devices, other electronic equipment, or software. However it has

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 39 / 62
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

to be clear that NFE, ESE and CHA will only be able to provide support on the basic level. More complicated technical questions or failures will be collected by the end-user help and will be transferred to the technical partners of ALFRED for further assistance.

In France the Netherlands and Germany the TP will receive an information sheet (in the workbook) with the contacts for support and questions. If necessary the helpdesk staff will provide support to the user via telephone or visit the older person at home to resolve any issues that older person is struggling with.

3.7.4 Workbook

In order to manage the TP within the pilots and support the researcher in following the research structure, each TP will have his own workbook in the different pilots and the goal of it is to provide continuous guidance for the piloting phase and to increase the test participant's self-confidence when interacting with ALFRED. This document is co-created by NFE, ESE and CHA who will then localize and adapt it for the country specific pilots. The workbook will contain the necessary surveys, interviews, guidance and ethical procedures. Once the information has been gathered (anonymously) the workbook will be destroyed in the end of the ALFRED project, and in consequence the informed consent will be a separated document from the workbook.

3.7.5 Overview Equipment

For the moment, Table 19 describes the testing material that is necessary for the pilot phase. This list will be updated and completed as the ALFRED system development moves forward in order to make sure that any necessary material has not been dismissed.

The main idea is that the each test person has at their disposal all the necessary testing material, notably the Android Smartphone on which the ALFRED system is running, throughout the testing phase. The ALFRED system requires an efficient internet data connection to be used correctly, as a result, this will also provide for each participant and it will be configured in each smartphone before the testing starts.

In order to get a comprehensive picture of the individual user's health, information on vital data like heart rate, breathing frequency and body temperature will be collected with a sensor package to obtained to get a clear picture on the health and fitness status, general wellbeing or improvement/worsening of specific physical parameters. This data will be collected with help of sensor packages and also collect data when users participate in serious games for prevention and rehabilitation. Serious games can help to increase physical abilities, social competence, hand-eye coordination, dexterity and creativity.

Table 19: Testing Material for the Pilot Phase

Pilot 1 (The Netherlands)	Pilot 2 (Germany)	Pilot 3 (France)
10 Android (Google) Smart Phones	20 Android Smart Phones	5 Android Smart Phones
5 Sensors packages	12 Sensor packages	4 sensor packages
10 flatrate internet data connections	40 flatrate Internet data connections	5 internet data connections (one for each phone)
5 serious game packages	2 Sensor packages	4 serious game packages

3.7.6 Ethical management process

The ethical management process is detailed in the D8.1.1 underlining that the end-users' rights and privacy will be safeguarded throughout the different project testing phases involving the representatives of the target groups. However, these ethical issues are essential and in consequence this chapter emphasizes some key points for the ethical methodologies.

In each different testing phase, the involved end-users are asked to fill in the informed consent (see example in the D8.1.1 Appendix 1) ensuring that the participation in the evaluation is voluntary. This procedure is done with each single participant and in each of the testing phases so that the participant signs the most relevant and recent version of the consent form.

Regarding the end-users' personal data protection, the test administrators will assure that all the personal data is handled anonymous and use the aggregated data for the analysis of the pilot testing results. All the collected personal data will be stocked in a safe manner and only in the test administrator's organization. Furthermore, the end-user organization carry out a continuous monitoring related to the relevant legislation on the European and national level. Finally, during the whole project and testing time, each test person has the right to withdraw from the project with no delay and without any cost. The study personnel and make sure that all the persons are informed about this possibility.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 41 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

4 Individual Pilot Descriptions

As described in the previous chapter3, the ALFRED pilots will combine several evaluation perspectives in all the three testing countries, which will enable a holistic system validation and assessment approach. This chapter focuses on detailing the specificities of the pilots in the Netherlands (Pilot 1), in Germany (Pilot 2) and in France (Pilot 3). These country and pilot specific evaluation details will be updated and completed (where necessary) in the forthcoming WP8 deliverables. This flexible approach will ensure that the pilots will use the most relevant evaluation methodologies covering all the ALFRED-system.

On top of the iterative evaluation process, the ALFRED system will be validated by the end-users in the three different pilots. The last iterative testing will be carried out around M24-M26 (as stated in the D8.1.2, chapter 2.2). The first pilot will start at M23 followed by the pilot 2 (M26) and finish with pilot 3 (M29). Following this, the pilots will start taking place from the M26 until the last project months.

4.1 Specifications for Pilot 1 – Individual Usability

As reflected in Figure 4, the Pilot 1 – Individual Usability pilot (T8.2) will first and foremost focus on the usability of the ALFRED system, during different iterative evaluations. In a final usability research the overall ALFRED system will be tested with 20 end users, before the actual pilots at the homes of older people, making sure that older end users can use the entire ALFRED system independently during the pilot.

The final pilot evaluation in the Netherlands will focus on the impact on the lives of older end users, the impact on their social lives and their capacity to perform activities of independent living at home. It will also help to define the economic impact to bring ALFRED to the market.

4.1.1 Evaluation perspectives in Pilot 1

Although the Dutch pilot has a large focus on the usability perspective, it will also be used to demonstrate and validate other aspects of the ALFRED system. In Pilot 1, the following evaluation perspectives are involved:

- **Usability perspective:** started already in M12 with the iterative testing process and will be continued throughout the project. The Dutch pilot has the lead on this perspective.
- **Economic perspective:** focusing on the purchasing aspects of the tested solution by the involved end-users after they have used the entire ALFRED system at home.
- **The perspective of users:** focusing on the independent usage of the solution by the involved end-users in their home environment.

Furthermore, the KPIs and the envisaged evaluation methods for each of these perspectives, which are described in the chapter3.2.2, will be implemented in Pilot 1.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 42 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

4.1.2 Involved Test Persons in Netherlands

Table 20 gives an overview of the involved test persons in the Dutch pilot. The TP's from the final usability research can also participate in the Dutch pilot, using the ALFRED solution independently at home. A total of 20 TP's will be involved in the final usability research and a total of 10 TP's will be involved in the final pilot at the homes of older people.

Table 20: Definition of the Involved Target Groups in Pilot 1

Involved Target User Group	Role in the Pilot 1
PTG1: Independently living and autonomous older adults 60+	<p>Participate in the final usability research.</p> <p>Test the complete ALFRED-system autonomously in the home environment and in a day-to-day usage for a predefined period of time.</p> <p>Participate in the system evaluation by using the different evaluation tools.</p>
PTG2: Independently living older adults 60+ having informal caregiver's (friend, family member, spouse) support	<p>Participate in the final usability research.</p> <p>Test the complete ALFRED-system autonomously and with the informal caregivers in their home environment and in a day-to-day usage for a predefined period of time.</p> <p>Participate in the system evaluation by using the different evaluation tools.</p>
PTG3: Independently living older adults 60+ having formal caregiver's support	<p>Participate in the final usability research.</p> <p>Test the complete ALFRED-system autonomously in their home environment and in a day-to-day usage for a predefined period of time.</p> <p>Participate in the system evaluation by using the different evaluation tools.</p> <p>In addition, with this target group, a special attention will be given to the ALFRED Pillar III and Pillar IV Apps that can improve their physical care.</p>
STG1: Informal caregivers (such as friends, family members of the older adults)	<p>This person is informal caregivers of one of the TPs that is involved in the Pilot 3.</p> <p>She/he will be given the ALFRED-solution so that she/he can interact with the person she/he is taking care of by using ALFRED.</p> <p>Test the complete ALFRED-system with the person that she/he is taking care of in the home environment and in a day-to-day usage for a predefined period of time.</p> <p>Participate in the system evaluation by using the different evaluation tools.</p>

4.1.3 Timeline and Pilot 1 Phases

The following table reflects the timeline of the Dutch pilot organization in further detail.

Table 21: Initial Time Planning for Pilot 1

Month	Pilot Task	Details
M12-M26	Iterative Testing Evaluation	This testing helps to prepare the actual pilot phase (define the most relevant evaluation tools, localize the ALFRED apps, get in touch with interested end-users etc.)
M23-M26	Final usability research with 20 end users, applying the learning method.	This will be the final usability research, on a larger scale to test all usability aspects of the overall ALFRED system.
M26-M28	Preparation of D8.2.1 and 8.2.2,	The deliverable will contain the results from the usability evaluations and the final pilot plan, including workbooks and all related material for the pilot.
M30	Preparation of the testing sites	Installation of the necessary testing material etc.
M31	First testing wave with 5 older adults at their home environment during 3 weeks and with at least one informal caregiver	In the end of these 3 weeks, each TPs is interviewed. In between the two testing waves there is at least one week so that the test administrator has time to address the various issues that might have been raised during the first piloting wave.
M32	Second testing wave with new 5 older adults at their home environment during 3 weeks and with at least one informal caregiver.	The participants are different from the first testing wave. In the end of these 3 weeks, each TPs is interviewed.
M33-36	Gathering of all the Pilot 1 data and start of the data analysis for D8.2.3 and D8.2.4.	The collection of the results and preparation of the analysis work.

4.1.4 Methodologies for Pilot 1

As explained in the previous section, Pilot 1 will perform research on the usability perspective, using the defined usability methodology in D8.1.1 in iterative evaluations. In the Netherlands the usability study will be performed in a controlled environment at the premises of the National Foundation for the Elderly.

The Dutch pilot will continue to use for the iterative usability evaluation the ALFRED workbook designed in D8.1.1. For the final usability evaluation the workbook will be updated in order to include all aspects of the ALFRED system.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 44 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

The final usability research will be a longitudinal usability research with 20 TP's on the overall ALFRED system. During this usability research all aspects of the ALFRED system will be tested with the TP. In order to make reliable predictions on the independent use of the ALFRED system for the pilot running at the homes of the TP's, the research will take the shape of a Longitudinal Usability Study. As for older people, the whole concept of talking to a system, takes a little getting used to, a Longitudinal Usability Study will make it possible to test the ALFRED system multiple times (2 times). This may help in detecting which problems that show up in the first session will disappear on itself and acquire no or little attention, compared to problems that require more re-design. In this phase of the project, many components of the ALFRED system have been tested, during the iterative evaluations and many design aspects will be more difficult to change. As we are working with the specific target group of older people, it is important to adapt the research to these circumstances. This final longitudinal usability research will make it possible to make more thorough conclusions about the usability for the specific target group of ALFRED.

In a second phase, the Dutch pilot will implement the pilot at the homes of 10 older people within the end user perspective. From the end user perspective, the Dutch pilot will implement the OPQL-35 combined with the TAM methodology [section3.5.1] in order to define the impact on life and activities of daily living and the added value of the ALFRED system on this. In D8.2.1. this pilot running and evaluation with 10 persons at their homes will be further defined. The workbooks will be set up to guide the research in practice. The pilot will be performed at the homes of older people (primary target groups), living in the province of Utrecht. Most of these people will not have an internet connection, the pilot will have to provide for a (mobile) internet connection.

Data will be gathered through a combination of log-in data, questionnaires and interviews.

4.2 Specifications for Pilot 2 – Hospital Environment

In Germany the pilots will be performed either within a hospital environment on the premises of the Evangelisches Geriatriezentrum (EGZB) or within the private homes of the ALFRED end users. The pilot will focus mainly on the physical wellness and prevention of older adults and also evaluate the socioeconomic and usability perspective in the context of physical wellness and prevention.

4.2.1 Evaluation perspectives in Pilot 2

The German pilot directs its attention mainly on the physical wellness and prevention aspect, however it will also evaluate the usability and the economic perspective of the ALFRED system. In Pilot 2, the following evaluation perspectives are involved:

- **Usability perspective:** started already in M12 with the iterative testing process and will be continued throughout the project.
- **Economic perspective:** focusing on the purchasing aspects of the tested solution by the involved end-users after they have used the entire ALFRED system at home.
- **Health perspective:** focuses on physical wellness and prevention related aspects like increased amount of physical activity or mental wellbeing.

Furthermore, the KPIs and the envisaged evaluation methods for each of these perspectives, which are described in the chapter 3.3.2, will be implemented in the Pilot 2.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 45 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

4.2.2 Involved Test Persons in Germany

Table 18 gives an overview of the involved test persons in the German pilot.

4.2.3 Timeline for Pilot 2

The iterative testing in Pilot which is ongoing since M12 will carry on testing with ALFRED users until M26. The following is the timeline of the German pilot organization in further detail.

Table 22: Initial Time Planning for Pilot 2

Month	Pilot Task	Details
M18-M26	Iterative Testing Evaluation	This testing helps to prepare the actual pilot phase (define the most relevant evaluation tools, get in touch with interested end-users etc.)
M23	D8.3.1 "Piloting and Validation Report II: Hospital"	Preparation of the Template
M23-M26	Pilot preparation	Finalisation of the pilot details to obtain ethical vote and data protection vote
M26-M27	Recruitment process	Recruiting test participants from PTG1, PTG2, PTG3, PTG4, STG1 and STG2
M26-M28	D8.3.1 "Piloting and Validation Report II: Hospital"	Preparing the deliverable
M27-M30	1 st test phase	Conducting the first test phase with 20 older adults in a healthcare environment
M29	D8.3.1 "Piloting and Validation Report II: Hospital"	Ready for internal review
M30	D8.3.1 "Piloting and Validation Report II: Hospital"	Submission
M30	D8.3.2 "Piloting and Validation Report II: Hospital"	Preparation of the Template
M31-M33	2 nd test phase	Conducting the second test phase with 20 older adults within their dwellings
M31-M35	D8.3.2 "Piloting and Validation Report II: Hospital"	Preparing the deliverable
M35	D8.3.2 "Piloting and Validation Report II: Hospital"	Ready for internal review
M36	D8.3.2 "Piloting and Validation Report II: Hospital"	Submission

4.2.4 Methodologies for Pilot 2

Pilot 2 will evaluate the physical wellness and prevention aspect of ALFRED.

Modern sensor technology will be used to assess directly measurable values like heart rate, breathing frequency and body temperature of the user. These values can be used to check for physical adaptations of the ALFRED system as a result of the exercise stimulus that will be provided in Pilot 2, with the help of serious games. Another method for the evaluation is the Body mass index (BMI). A reduction of the BMI has a positive and preventive effect on diseases like high blood pressure, osteoarthritis, diabetes, cardiovascular disease and some cancers. Further methods to determine the physical wellness and prevention effects of ALFRED on the older adults will be measured with questionnaires and medical assessments. Examples for these assessments are the BBS for the evaluation of balance or the FTSST test, measurement of grip strength, the 6 minute walking test and the overall measurement of physical activity of older adults. A comprehensive list of assessments that will provide information on the impact of ALFRED to the physical wellbeing has been defined in D6.1. Mental and cognitive influence of ALFRED of the users will be measured with the Mini Mental State examination Test (MMSE) or the Warwick-Edinburgh Mental Well-being scale. The PSQI is a tool to assess the sleep quantity and therefore allows a combination of physical and mental evaluation of the effects of ALFRED to the older adults.

4.3 Specifications for Pilot 3 – Day-to-day usage

The Pilot 3 is managed by E-Seniors Association in France, this pilot planning started in the D8.1.2 and now it is updated according to the pilot definitions specified by the end-user partners. This Pilot is the last pilot, starting in the M29, and as a result its main goal is to evaluate the entire ALFRED system with the primary and secondary end-users in day-to-day usage. The final results and recommendations resulted from this pilot are reported in the deliverable 8.3 “Piloting and Validation Report III: Association” M36.

4.3.1 Evaluation Perspectives in the Pilot 3

ESE works hand in hand with NFE and CHA to design the most adapted evaluation tools and the aim of this is to employ evaluation methods that are harmonious in the piloting countries that use the same evaluation perspectives (that are explained in the chapter 3.1).

In the Pilot 3, the following evaluation perspectives are involved:

- **Usability perspective:** started already in M12 with the iterative testing process and will be continued throughout the project.
- **Economic perspective:** focusing on the purchasing aspects of the tested solution by the involved end-users.
- **The perspective of users:** focusing the independent usage of the solution by the involved end-users in their home environment.

Furthermore, the KPIs and the envisaged evaluation methods for each of these perspectives, which are described in the chapter 3.5, will be implemented in Pilot 3.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 47 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

4.3.2 Involved Test Persons in France

Table 23 details more explicitly these specific target groups that are included in Pilot 3.

Table 23: Definition of the Involved Target Groups in the Pilot 3

Involved Target User Group	Role in the Pilot 3
PTG1: Independently living and autonomous older adults 60+	<p>Test the complete ALFRED-system autonomously in the home environment and in a day-to-day usage for a predefined period of time.</p> <p>Participate in the system evaluation by using the different evaluation tools.</p>
PTG2: Independently living older adults 60+ having informal caregiver's (friend, family member, spouse) support	<p>Test the complete ALFRED-system autonomously and with the informal caregivers in their home environment and in a day-to-day usage for a predefined period of time.</p> <p>Participate in the system evaluation by using the different evaluation tools.</p>
PTG3: Independently living older adults 60+ having formal caregiver's support	<p>Test the complete ALFRED-system autonomously in their home environment and in a day-to-day usage for a predefined period of time.</p> <p>Participate in the system evaluation by using the different evaluation tools.</p> <p>In addition, with this target group, a special attention will be given to the ALFRED Pillar III and Pillar IV Apps that can improve their physical care.</p>
STG1: Informal caregivers (such as friends, family members of the older adults)	<p>This person is informal caregivers of one of the TPs that is involved in the Pilot 3.</p> <p>She/he will be given the ALFRED-solution so that she/he can interact with the person she/he is taking care of by using ALFRED.</p> <p>Test the complete ALFRED-system with the person that she/he is taking care of in the home environment and in a day-to-day usage for a predefined period of time.</p> <p>Participate in the system evaluation by using the different evaluation tools.</p>

It should be made clear that the formal caregivers (STG2) are not directly included in this pilot as the testing is taking place in the TPs home environment. Also, the STG2 is involved in the Pilot 2 that will enable to collect their feedback as well.

In total, the Pilot 3 will involve 8 TPs from the primary target groups (PTG) and at least 2 informal caregivers (STG1) of the older adults involved in the Pilot 3.

4.3.3 Timeline and the Pilot 3 Phases

The very first step towards the Pilot 3 is the iterative testing process that is carried out by the three end-users partners from M12 onward. This additional testing process enables on the one hand to acquire end-user feedback from the three countries from the beginning of the ALFRED prototype development and on the other hand, to plan more efficiently the final pilots, and notably related to the design the pilot evaluation tools.

The recruitment of the TPs will be carried out in the very beginning of Pilot 3 (M29). The persons who participated in the iterative testing sessions in France will be invited to take part in the pilot (if they are interested). ESE's member and partner network will be used to find the rest of the TPs.

Then, once all the TPs are recruited, the preparation of the testing sites (in the TP home environment) starts ensuring that they get all the necessary testing material in time and that they have well understood their engagement in the pilot process.

The actual field testing will be done in two different waves between M31 and M32 with two different groups of 5 end-users. This pilot in two phases enables the test administrator to coordinate better the piloting process and be well available throughout the pilot testing in the individual homes. In addition, in between these two testing waves there should be at least one week so the Test Administrator (TA) can address the various issues (technical, ethical etc.) might potentially have been raised during the first piloting wave.

After the individual testing (M33), all the participants will be gathered in focus group discussions with the aim to enable the TPs exchanges their usage experiences with ALFRED together and to gather final qualitative feedback for the users.

The results are recommendations of the Pilot 3 and will be reported in the D8.3 that is due in M36. This report will be written throughout the T8.3 (M29 – M36). Table 24 below gives the initial time planning for the ALFRED pilot that is carried out in France. This time planning can be updated and completed if necessary.

Table 24: Initial Time Planning for Pilot 3

Month	Pilot Task	Details
M12-M26	Iterative Testing Evaluation	This testing helps to prepare the actual pilot phase (define the most relevant evaluation tools, get in touch with interested end-users etc.)
M29	Recruitment of the test participants	Participants from PTG1, PTG2, PTG3 and STG1
M30	Preparation of the testing sites	Installation of the necessary testing material etc.
M31	First testing wave with 4 older adults at their home environment during 3 weeks and with at least one informal caregiver	In the end of these 3 weeks, each TPs is interviewed. In between the two testing waves there is at least one week so that the test administrator has time to address the various issues that might have been raised during the first piloting wave.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 49 / 62
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

M32	Second testing wave with new 4 older adults at their home environment during 3 weeks and with at least one informal caregiver	The participants are different from the first testing wave. In the end of these 3 weeks, each TP is interviewed.
M33	Organization of two separate focus group sessions in each 5 participants	Animated by at least two test administrators.
M34	Gathering all the Pilot 3 data and the begging of the data analyse for the D8.3	The collection of the results and preparation of the analysis work.
M35 – M36	Writing the Deliverable 8.3: Piloting and Validation Report III: Association	The evaluation results are gathered and analysed aside the pilot performing
M36	Deliverable 8.3: Piloting and Validation Report III: Association	The report is submitted

4.3.4 Methodologies for Pilot 3

The key is to make sure that the TPs are using ALFRED daily, and for this two tools are designed:

- Diary
- System Activity Log (created by the technical project partners)

In addition to this, the evaluation methods defined in the usability, socio economic and user perspectives (see chapter3.1) are implemented. All these evaluation tools will be gathered in the Pilot Workbook that is given for the user. The test administrator (TA) is in charge of verifying that the TP completes the workbook correctly and for this reason the TA will visit or call (as the participant prefers) the TP at least once a week.

In the end of each individual testing phase (after the 3 weeks), each TPs is interviewed in order to check that all the evaluation material is correctly filled and to revise all the open questions related to this material. In the end of the interview, the TA gathers the evaluation material for the final analysis.

In addition to the evaluation tools used throughout the pilot phase at the older adult's home environment, two focus group discussion sessions will be organized. This way the TPs can share their testing experiences with each other and it enables the test administrator to collect final qualitative feedback from the participants.

Once the Pilot 2 in Germany, where the primary and secondary users use ALFRED in a controlled environment, has confirmed that the ALFRED-system can be used safely and smoothly in the individual home environment, the pilot focusing in the perspective of the users will start. The perspective of the users will be tested in the final pilot with the entire ALFRED solution in the Netherlands and in France. The results from the previous user evaluation sessions (from the iterative evaluation process and the clinical sessions) will be used to run this final pilot at the user's home environment. Before starting this pilot the usability and ease of use have been maximized and we know which apps can be used safely and independently by older end users for health and wellbeing purposes at home. The perspective of the users will then reflect an impact on the primary and the secondary target users themselves and the actual improvement in independent living at home.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 50 / 62
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

5 Conclusion

The iterative evaluations will continue in all three end-user countries until M26. The results of the pre-prototype evaluations that assessed the speech interaction of six tasks with twelve older adults are presented in the present deliverable. The test helped to gather information about how the users speak to ALFRED, enabled the evaluation of test-to-speech as well as the evaluation of the suitability of the selected utterances by detecting the test participants ability to understand what ALFRED said and meant. Since this was an early prototype evaluation the tests were performed with the simulated WoZ technique.

The first pre-prototypical evaluation of ALFRED was able to identify important issues and problems when handling the system. The issues that were raised in this deliverable now should be implemented into the system, before the next iterative testing cycle can start. The consideration of these points should allow a higher end user satisfaction of ALFRED after the next iterative testing. Despite of the high satisfaction values (ASQ and PSSOQ) with the current version of ALFRED it has to be mentioned that the current tests are based on a simulated WoZ experiment. It is not clear if the speech recognition software that is currently under development will have similar capabilities the simulated WoZ experiment.

The iterative tests will be continued in the will be conducted in M18-M21 (first prototype evaluation) and the Second prototype evaluations in M24-M26. These evaluations will involve further tests of the ALFRED system and report the results in the upcoming deliverables.

The pilot planning and the evaluation methods will be updated according to the feedback and reactions given by the project partners and the Advisory Board of the project. This flexibility ensures that the ALFRED pilots are as relevant and as efficient as possible.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 51 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

References

- [Dav89] Davis, F.D. "Perceived usefulness, perceived ease of use and user acceptance of Information Technology" *MIS Quarterly*, v13 no 3, 319-340., 1989.
- [HD+06] Halcomb E., Davidson P, Is verbatim transcription of interview data always necessary, *Applied Nursing Research* 19, P.38–42, 2006.
- [KH86] B. Klaus and P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.
- [Law69] Lawton, M.P. and Brody, E.M., "Assessment of older people: Self-maintaining and instrumental activities of daily living" *The Gerontologist*, v 9 No 3, 179-186., 1969.
- [Lew95] James R. Lewis, IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use, *International Journal of Human-Computer Interaction*, v.7 n.1, p.57-78, Jan.-March 1995
- [Lew91] Lewis, J.R., 1991. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. In: *ACM SIGCHI Bulletin*, vol. 23, pp. 78–81.
- [Nie94] J Nielsen, R.L Mack. *Usability Inspection Methods* Wiley, New York (1994)
- [Ost08] Osterwalder A, Pigneur Y (2010) *Business model canvas*. Wiley, Hoboken
- [Tud et al 11] C. Tudor-Locke, C. L. Craig, W. J. Brown, S. A. Clemes, K. De Cocker, B. Giles-Corti, Y. Hatano, S. Inoue, S. M. Matsudo, N. Mutrie, J.-M. Oppert, D. A. Rowe, M. D. Schmidt, G. M. Schofield, J. C. Spence, P. J. Teixeira, M. A. Tully, and S. N. Blair, "How many steps/day are enough? For adults," *Int J Behav Nutr Phys Act*, vol. 8, p. 79, 2011.

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 52 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

Annex A – ASQ Questionnaire

Overall, I am satisfied with the ease of use of the application for completing this task.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

Overall, I am satisfied with the amount of time it took me to perform this task with the application.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 53 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

Overall, I am satisfied with the supporting information (help function, on-screen messages and other documentation. NOT the instructor!) of the application for completing this task.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

Comments:

Please write down any comments about this task and/or the system below

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 54 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

Annex B – PSSUQ Questionnaire

1. Overall, I am satisfied with how easy it is to use this system.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

2. It was simple to use this system.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

3. I was able to complete the tasks and scenarios quickly using this system.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 55 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

4. I felt comfortable using this system.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

5. It was easy to learn to use this system.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

6. I believe I could become productive quickly using this system.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

(see next page!)

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 56 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

7. The system gave error messages that clearly told me how to fix problems.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

8. Whenever I made a mistake using the system, I could recover easily and quickly.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

9. The information (such as online help, on-screen messages and other documentation) provided with this system was clear.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

(see next page!)

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 57 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

10. It was easy to find the information I needed.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

11. The information was effective in helping me complete the tasks and scenarios.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

12. The organization of information by voice interaction was clear.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

(see next page!)

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 58 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

13. The interface (voice interaction) of this system was pleasant.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

14. I liked using the voice interaction of this system.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

15. This system has all the functions and capabilities I expect it to have.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

(see next page!)

D8.1.2 – Piloting Definition	Document Version: 1.0	Date: 2015-03-31	Status: For Approval	Page: 59 / 62
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

16. Overall, I am satisfied with this system.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

17. The vocal instructions were understandable.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

18. The audio volume was sufficient.

- I completely agree
- I agree
- I agree a bit
- Neutral/no opinion
- I disagree a bit
- I disagree
- I completely disagree

This is the end of the questionnaire. Thank you for helping us!

Annex D – PSSUQ Ratings

Participant ID	PS SU Q-1	PS SU Q-2	PS SU Q-3	PS SU Q-4	PS SU Q-5	PS SU Q-6	PS SU Q-7	PS SU Q-8	PSS UQ -9	PSS UQ -10	PSS UQ -11	PSS UQ -12	PSS UQ -13	PSS UQ -14	PSS UQ -15	PSS UQ -16	PSS UQ -17	PSS UQ -18
1	3	2	3	3	3	2	4	4	4	4	4	3	3	2	2	4	3	3
2	2	2	1	1	2	2	1	2	1	1	2	1	3	2	3	2	2	5
3	1	1	1	1	1	1	1	1	4	4	4	4	1	1	4	1	1	5
4	2	2	2	3	2	2	2	2	3	2	3	2	2	2	2	2	2	2
5	2	1	2	1	1	1	4	3	2	1	2	1	1	1	2	2	2	2
6	2	2	2	1	1	2	7	7	7	7	7	4	3	3	3	3	2	2
7	1	1	1	1	1	1	1	1	1	1	2	2	2	1	2	1	2	1
8	1	1	2	1	2	1	3	3	3	3	1	1	1	1	1	1	1	1
9	1	1	1	1	1	2	0	0	0	0	0	0	2	2	2	2	2	2
10	1	1	2	1	1	1	2	2	2	2	2	1	2	1	3	2	1	1
11	2	2	2	2	2	2	5	2	2	2	2	2	2	2	5	2	2	6
12	1	1	1	1	1	1	4	4	4	4	4	1	1	1	1	1	1	1