Project no.:
AAL-2009-2-137

PeerAssist

**A P2P platform supporting virtual communities to
assist independent living of senior citizens**

WP6 Deliverable 6.3.
"Final evaluation report"

Lead Participant/Editor	Ingema
Authors	Blanca Morales, Iker Laskibar, Leticia Coello, Foivos Demertzis, Orestis Panos

Table of Contents

Table of tables

Table of Figures

1. EVALUATION PROCEDURE

1.1. Introduction

The evaluation procedure for the project included trials in Spain and Greece, taking place in different times and using different versions of the system. Forty users were involved in total; 19 in Spain and 21 in Greece. In Spain, users were recruited from the groups of elderly people that Ingema works with. In Greece, users were recruited from the elderly friendship clubs that AEDA has direct access to through the Municipality of Athens. The PeerAssist system was prepared and complemented with various options and features, most of these options being from deliverable D2.2. (PeerAssist use scenarios definition) and from deliverable D2.3. (Definition of global requirements, trials and demonstration), in which the user's preferences and needs were collected.

In Spain, the methodology for the trials, was developed so as to take place in pairs. More specifically, there was a user and one evaluator in one room and another user with the another evaluator in a different room. Once users arrived to the lab where the trials were conducted, they were instructed to use the system by following a concrete route designed by the PeerAssist consortium. During the evaluations, the users were guided by an assisting person that was physically present to make them feel more comfortable during the trials. After the first trials in Spain, a second date was arranged for the next week in order to let them test the PeerAssist system again and report any differences in their perception of the PeerAssist system between the two trials.

The trials in Greece followed those in Spain, taking advantage of the expertise and overall evaluation obtained there, and using an improved user interface that was designed to address the users' comments from the Spanish trials. The evaluation methodology in Greece included both pairs of users that tried the system at AEDA premises, but also users that were communicating through PeerAssist from their houses, experiencing the same set of scenarios from their own comfortable environment, to allow more realistic experience. Some photographs of the trials are included below.
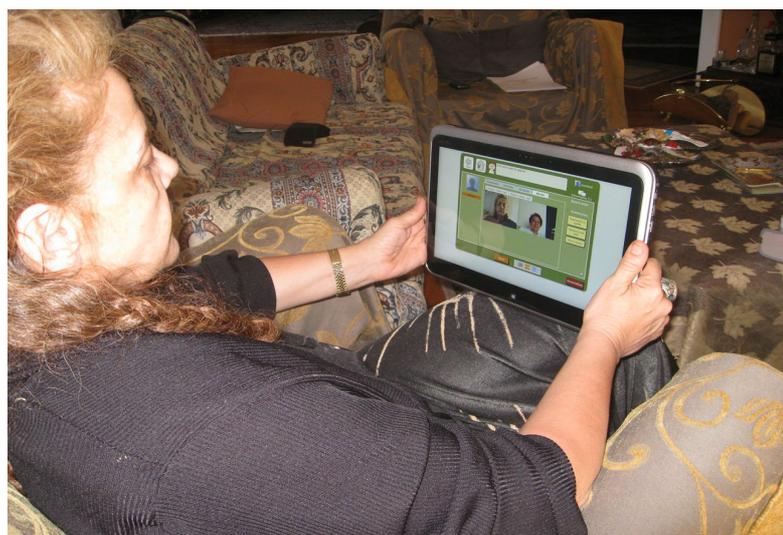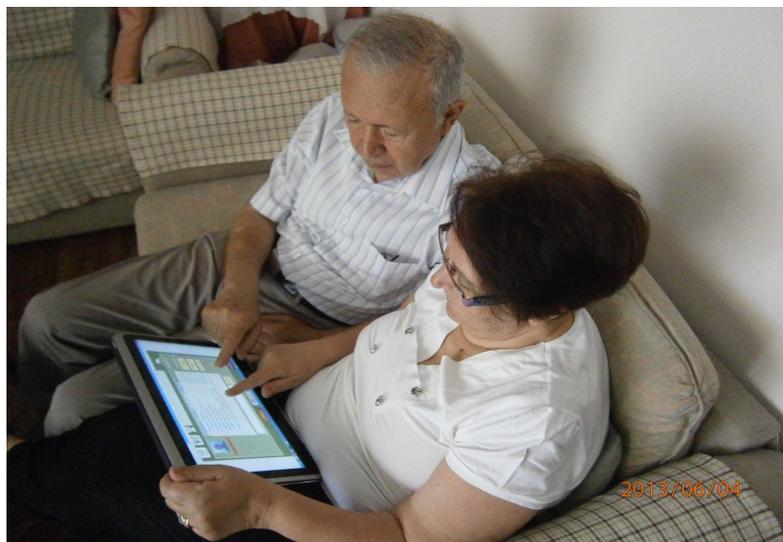
Figure 1: Photos from the PeerAssist trials

In both sites, the following features of PeerAssist system were evaluated:

Feature	Description
Usability	How **easy** and **pleasant** are these features to use. This is a **quality attribute** that assesses how easy user interfaces are to use. The word "usability" also refers to methods for improving ease-of-use during the design process. It is also the functional evaluation of the capacities of the system: usefulness related to the designed capabilities of the system, the new alternative capabilities the user can be benefited from and the indirect synergistic relations between the user and the system interaction.
Acceptability	The response related to the acceptance, tolerance and how much a user likes the system he/she tries. It includes personal likeness, trustworthiness reliability etc. Two factors influence the decision: **perceived usefulness** and **perceived ease-of-use.**
Accessibility	This is the degree to which a product, device, service, or environment is **available** to as many people as possible. It describes its ease of reach, use and understanding. In terms of user experience design it can also be related to the overall comprehensibility of the information and features presented to a user. It contributes to shorten the learning curve attached with the system. The term is more related to structures and architectural barriers, could be applied as the simplicity to use, and the e-inclusion efficiency that a system could develop compared to a different system with the same functionalities.
Time to complete the tasks	The time a user spends to complete specific tasks. Naturally, this influences the other features above.

The study was carried out in 9 steps:

1. Introduction and informed consent
2. Sociodemographic data
3. Scenario 1 "User profile" – After-Scenario Questionnaire (ASQ)
4. Scenario 2 "Social profile" – After-Scenario Questionnaire (ASQ)
5. Scenario 3 "Search service" – After-Scenario Questionnaire (ASQ)
6. Scenario 4 "Caregivers" – After-Scenario Questionnaire (ASQ)
7. Scenario 5 "Notifications" – After-Scenario Questionnaire (ASQ)
8. Post-Study System Usability Questionnaire (PSSUQ)
9. Closing

1.2. User profile

Regarding the user profile, initially the inclusion criteria to participate in the evaluation was, older adults over 60 years old, without cognitive impairment, in good physical and mental conditions, living independently and being active. On the other hand, taking into account the essence of the PeerAssist project, it was not necessary that the users have been familiarized with the new technologies.

User profile	Older adult (+60 years old)
	Without cognitive impairment
	Good physically
	Good mental conditions
	Living independently
	Active life style

1.3. Protocol and data collection plan

The users carried out some tasks, which were developed based on the scenarios of deliverable D2.2. (PeerAssist use scenarios definition) and explained in detail in deliverable D6.1 (Evaluation design report). During the evaluation, each user was accompanied by one assistant. The main functions of that assistant was:

- Introduce details about the platform.

- Explain the consent form and possible doubts or questions.

- Help if they are not able to perform the tasks after trying to do it.

- Observe users behavior.

- Administer questionnaires.

Regarding the collecting data, there are different methodologies available to proceed. One of these classifications is depending on the degree of interaction between the users and the assistants. According to this classification, there is a division of three categories: observation only, interacting with the user and method supplements. It is very important to select the method that will best address the goals of our study, as well as resources available to conduct the evaluations.

In our case, due to the necessity of interaction between the user and the assistant, we chose the mixed option of observation and ad-hoc interview with validated questionnaires. One of the main advantages of this mixed methodology is the opportunity to see the user interacting with the system and in some concrete moments to ask a question regarding her/his execution.

1.4. Evaluation of scenarios

All the users followed the same route in order to be able to interact with other participants. The route can be seen below:

User profile

Social profile

Search service

Caregivers

Notification
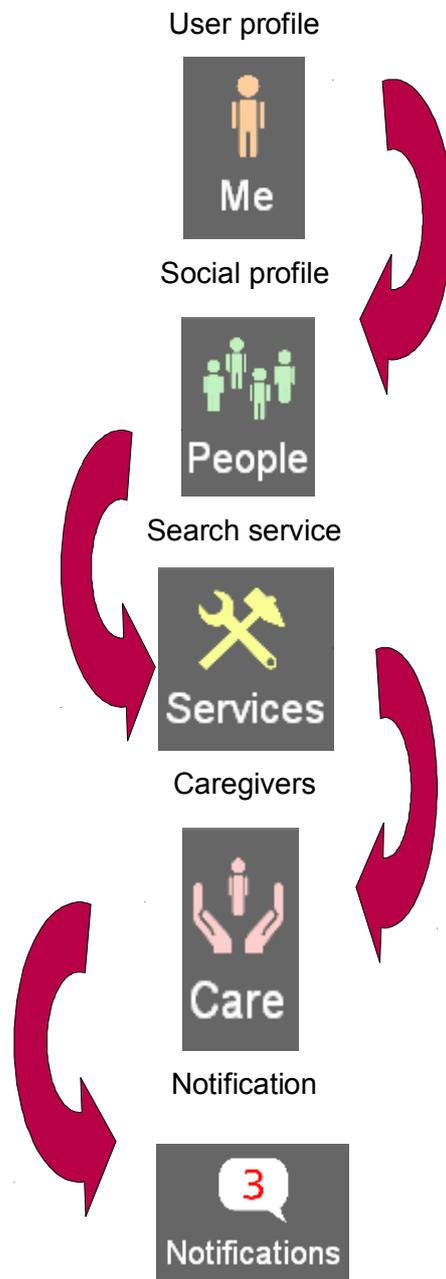
Figure 2: Evaluation process

2. GENERAL RESULTS

2.1. Sociodemographic data in Spain

Spanish users were people from 60 to 79 years old (mean=70.11, sd=5.17). The following tables show the distribution of the participants taking into account their sex, the type of residence were they live, their academic degree and their technology usage experience.
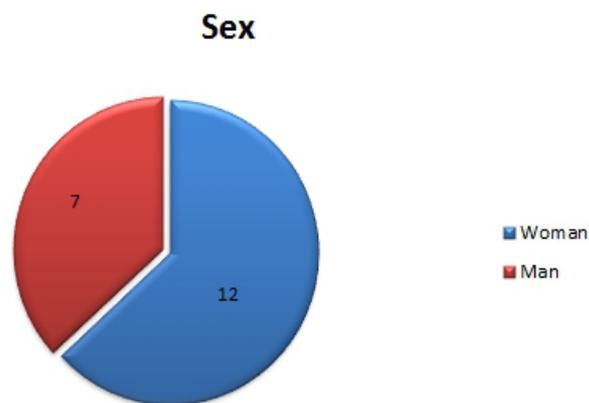
Figure 3: Participants by sex
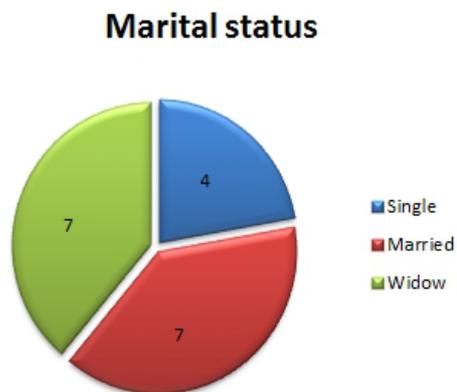
Figure 4: Participants by marital status

Type of residence
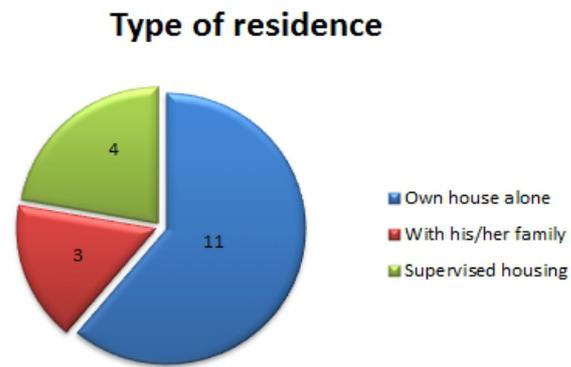
Figure 5: Participants by residence type

Highest academic degree
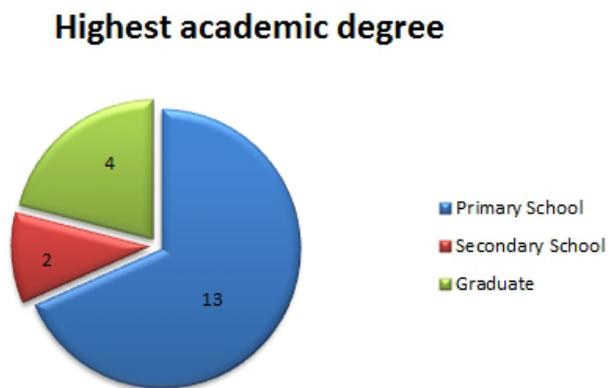
Figure 6: Participants by academic degree

How often do you use a computer?
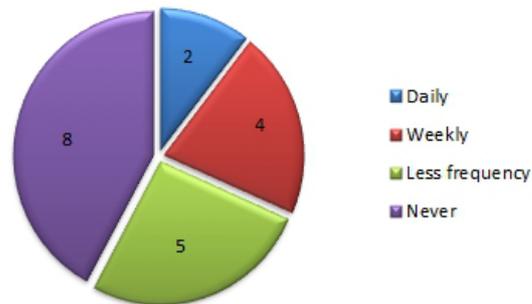
Figure 7: Participants by computer usage

How often do you use the internet?
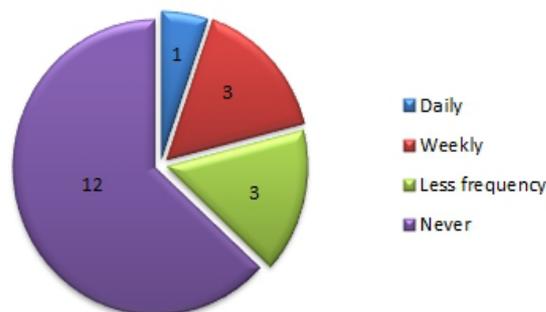
Figure 8: Participants by internet usage

2.2 Sociodemographic data in Greece

Greek users were people from 63 to 82 years old (mean=67.45, sd=5). The following tables show the distribution of the participants taking into account their sex, the type of residence were they

live, their academic degree or their technology usage.

Sex
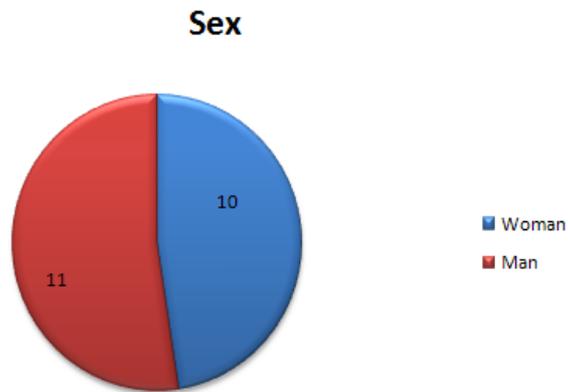
Figure 9: Participants by sex
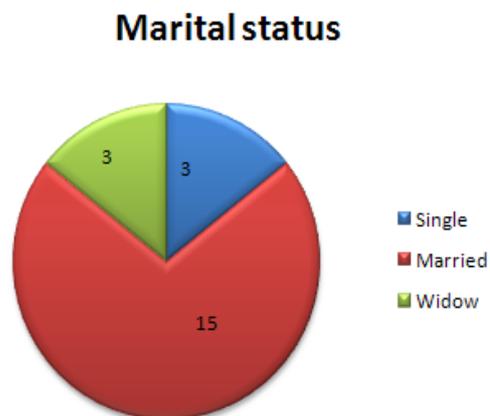
Marital status

Figure 10: Participants by marital status

Type of residence
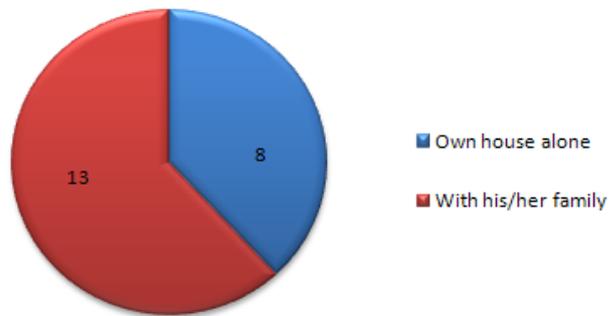
Figure 11: Participants by residence type

Highest academic degree
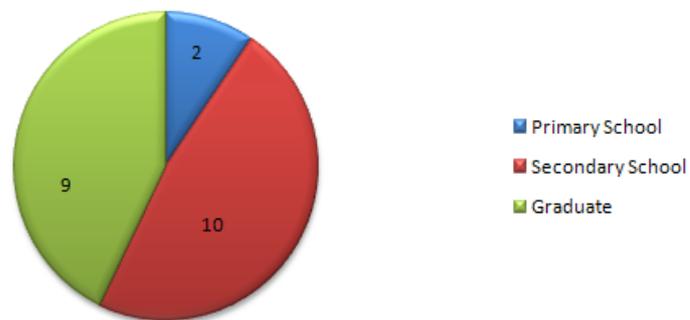
Figure 12: Participants by academic degree

How often do you use a computer?
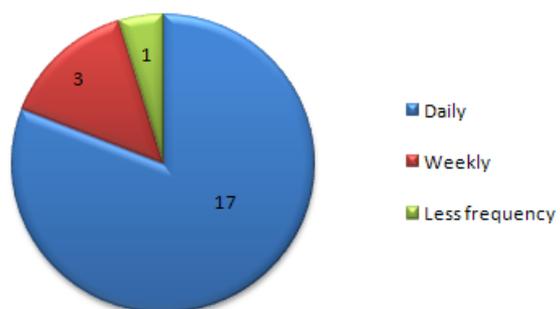
Figure 13: Participants by computer usage

How often do you use the internet?
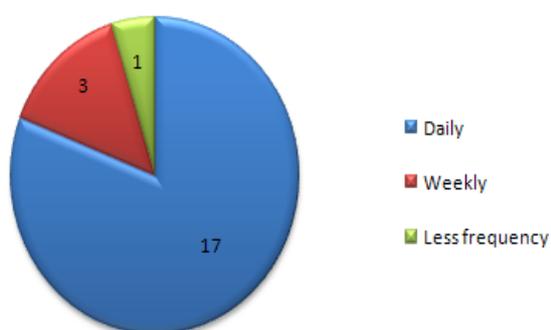
Figure 14: Participants by internet usage

2.3. In-deep analysis of the scenarios

2.3.1. Scenario 1: User profile

First of all, the users accessed the main menu and got familiar with the interface of the PeerAssist system. The assistant explained the general idea of the PeerAssist system and showed them the different options that they could find there. In the first scenario, they were asked to work on the option regarding the user profile. For that purpose, users had to click the "Me" option (or "Profile" in the Greek version) and change some information about their profile such as their name, surname, date of birth, sex, hobbies, etc. (Task 1). After that, they had to access the blog and write their first message there (Task 2). As this trial was done in pairs, the main idea was to share this message with the other user.

2.3.1.1. Spanish results for scenario 1

Since the users were asked to carry out the same tasks in the first and in the second round of this trial, the results of the tasks were compared by the Wilcoxon test [1] in order to know if the

differences found between the first and second round were statistically different. Results are presented in the following tables and figures. Just to mention that the items are based on the 7-point Likert scale, ranged from "Strongly agree" with value 1 to "Strongly disagree" with value 7.

Task 1 and 2

		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this scenario?	Pre	19	1.63	0.761	0	3	0.020	2.332
	Post	19	2.16	0.688	1	3		
1.2. The user is able to complete the general scenario?	Pre	19	1.16	0.375	1	2	0.157	-1.414
	Post	19	1.26	0.452	1	2		
1.3. Has the user completed the two activities of the scenario 1?	Pre	19	1.79	0.631	1	3	0.059	-1.890
	Post	18	2.11	0.471	1	3		
2.1. Does the user feel comfortable using the system during the scenario 1?	Pre	19	2.16	0.765	1	3	0.008	-2.640
	Post	19	2.68	0.478	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	19	2.47	0.964	0	3	0.197	-1.289
	Post	19	2.74	0.562	1	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	19	2.68	0.820	0	3	1	0.000
	Post	19	2.74	0.452	2	3		
3.1. Do you like the idea of this scenario?	Pre	15	2.07	0.799	1	3	0.157	-1.414
	Post	17	2.35	0.702	1	3		
3.2. Do you like the PeerAssist idea if you future situation changes?	Pre	18	1.78	1.003	0	3	0.776	-0.284
	Post	18	1.72	1.179	0	3		
3.3. Will you buy this system if it was in the market?	Pre	18	0.89	1.079	0	3	0.719	-0.360
	Post	19	0.74	0.991	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	18	1.44	1.097	0	3	0.835	-0.208
	Post	19	1.32	1.293	0	3		
3.5. Does the user like better other communication device?	Pre	12	0.17	0.389	0	1	0.157	-1.414
	Post	11	0.45	0.522	0	1		
1. Overall. I am satisfied with how easy of completing the tasks in this scenario	Pre	19	3.05	1.545	1	7	0.923	-0.096
	Post	17	3.06	1.638	1	6		
2. Overall. I am satisfied with the amount of time it took to complete the tasks in this scenario	Pre	19	3.53	1.467	1	6	0.724	-0.353
	Post	17	3.29	1.611	1	6		
3. Overall. I am satisfied with the support information (online-line help, messages, documentation, and training material when completing the tasks.	Pre	19	2.89	1.696	1	7	0.726	-0.351
	Post	17	2.59	1.543	1	5		

Table 1: Results of the 1st scenario (Task 1 and 2)
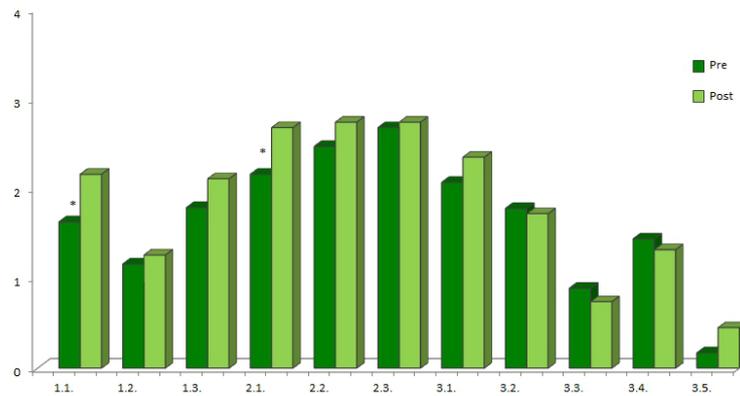
Figure 15: Results of the 1st scenario
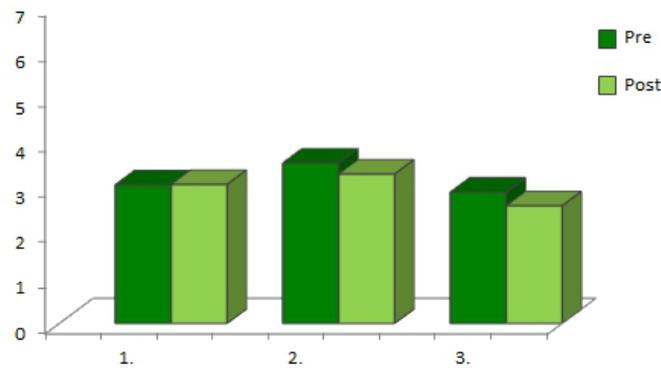
Figure 16: ASQ results of the 1st scenario

As it has been shown in the previous table, some significant differences have been found. Users interacted much better with this scenario the second time than the first one (Z= 2.332, p<0.05). On the other hand, users felt more comfortable in the second trial than in the first one (Z=-2.640, p<0.05). No differences have been found in the ASQ results. The best rated items were the ones regarding how comfortable they felt, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it. The score of 3.5 is considered low because of the scale (1=best, 7=worst).

2.3.1.2. Greek results for scenario 1

Results from the Greek trials are presented in the following figure:

1.Usability / 2.Time / 3.Support
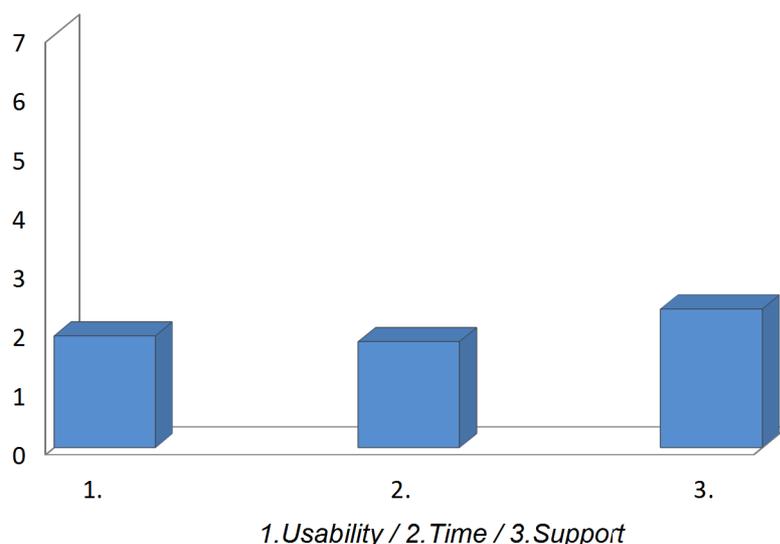
Figure 17: ASQ results of the 1st scenario

Likert test (1-7 scale) showing usability, time taken to complete the task and accesibility by the provided information that supports the system. The best rated items were the ones regarding how comfortable the feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it similarly to the Spanish trials. Overall, the results for this scenarios in Greece were considerably better than in Spain, probably due to the improved user interface used in this case.

2.3.2. Scenario 2: Social profile

This scenario includes task 3 to task 8 from the evaluation design explain in deliverable D6.1:

• Task 3 consisted of searching in the PeerAssist system for the other user (the user executing the trial at the same time), send him/her an invitation to be a new friend or accept the invitation done by the other user.

• Task 4 consisted of reading the information of the other user (name, date of birth, hobbies). They also had to read what the other user had written in the blog.

• Task 5 consisted of using the chat with the other user. Normally, they started writing general issues, but users were slowly gaining confidence and writing more and more.

15

• Task 6 consisted of initiating or accepting a videocall. They used to spend about 5 minutes speaking and seeing each other using this system.

• Task 7 consisted of creating a new group in order to share hobbies. For that purpose, one of the users had to create a new group, and share it with the other user who had to accept the invitation and be member of that group.

• Task 8 consisted of chatting using the group created on task 7 (group chat).

2.3.2.1. Spanish results for scenario 2

Task 3

		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this general task 3?	Pre	18	2.44	0.616	1	3	0.763	-0.302
	Post	16	2.25	0.775	1	3		
1.2. The user is able to complete the general task 3?	Pre	18	1.50	0.514	1	2	1	0.000
	Post	16	1.44	0.512	1	2		
1.3. Has the user completed the activities of the general task 3?	Pre	18	2.56	0.616	1	3	0.083	-1.732
	Post	16	2.13	0.619	1	3		
2.1. Does the user feel comfortable using the system during the general task 3?	Pre	18	2.44	0.784	1	3	0.725	-0.351
	Post	17	2.47	0.514	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	18	2.44	0.922	1	3	0.102	-1.633
	Post	17	2.59	0.795	1	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	18	2.50	0.857	1	3	0.129	-1.518
	Post	17	2.65	0.702	1	3		
3.1. Do you like the idea of this general task 3?	Pre	13	1.85	0.987	0	3	0.705	-0.378
	Post	14	2.00	1.038	0	3		
3.2. Do you like the general task 3 idea if you future situation changes?	Pre	16	1.75	1.125	0	3	0.792	-0.264
	Post	16	1.81	1.223	0	3		
3.3. Will you buy this concrete system option if it was in the market?	Pre	16	1.00	1.095	0	3	0.763	-0.302
	Post	16	0.94	1.124	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	15	1.67	1.175	0	3	0.236	-1.186
	Post	15	1.67	1.234	0	3		
3.5. Does the user like better other communication device?	Pre	10	0.60	0.516	0	1	1	0.000
	Post	7	0.86	0.378	0	1		

Table 2: Results of the 2nd scenario, task 3
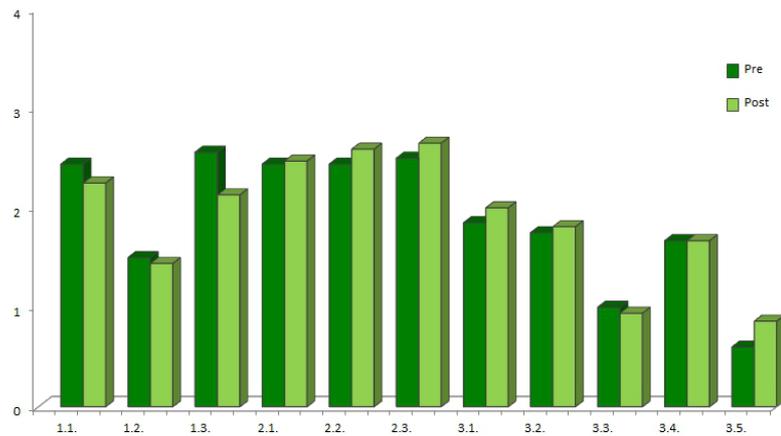
Figure 18: Results of the 2nd scenario, task 3

Spanish users had a good score at this task. Nevertheless, their opinion didn't change a lot between the first and second trial . The best rated items were the ones regarding the easiness to complete the scenarios, how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it.

Task 4

		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this general task 4?	Pre	19	2.63	0.597	1	3	1	0.000
	Post	19	2.63	0.597	1	3		
1.2. The user is able to complete the general task 4?	Pre	19	1.68	0.478	1	2	0.655	-0.447
	Post	19	1.63	0.496	1	2		
1.3. Has the user completed the activities of the general task 4?	Pre	18	2.44	0.616	1	3	0.739	-0.333
	Post	19	2.53	0.612	1	3		
2.1. Does the user feel comfortable using the system during the general task 4?	Pre	19	2.63	0.496	2	3	0.317	-1.000
	Post	19	2.74	0.452	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	19	2.63	0.761	1	3	0.098	-1.656
	Post	19	2.95	0.229	2	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	19	2.74	0.653	1	3	0.157	-1.414
	Post	19	2.95	0.229	2	3		
3.1. Do you like the idea of this general task 4?	Pre	17	1.94	0.827	0	3	0.270	-1.103
	Post	16	2.31	0.946	0	3		
3.2. Do you like this general task 4 idea if you future situation changes?	Pre	18	1.78	1.003	0	3	1	0.000
	Post	18	1.72	1.227	0	3		
3.3. Will you buy this this concrete system option if it was in the market?	Pre	18	0.94	1.110	0	3	0.679	-0.415
	Post	18	0.78	1.060	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	18	1.56	1.294	0	3	0.829	-0.216
	Post	18	1.44	1.247	0	3		
3.5. Does the user like better other communication device?	Pre	11	0.45	0.522	0	1	0.317	1.000
	Post	10	0.70	0.483	0	1		

Table 3: Results of the 2nd scenario, task 4
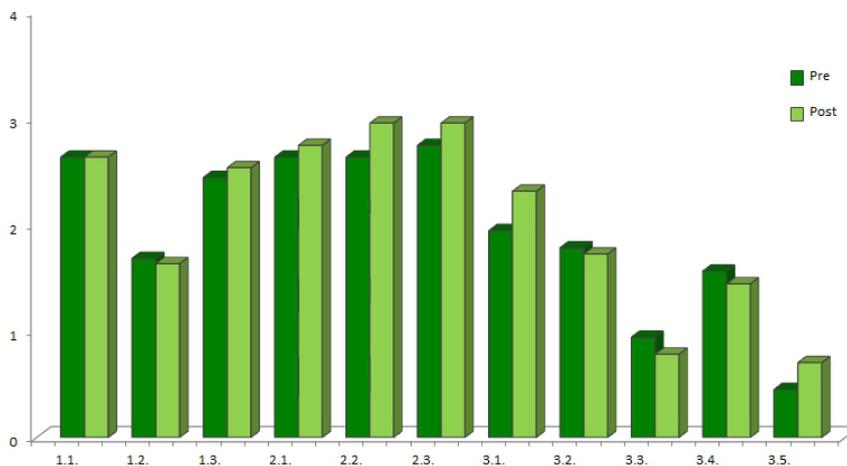
Figure 19: Results of the 2nd scenario, task 4

18

As with the previous task, the scores are high and the ones of the second trial tend to be better than the first ones, but no significant difference has been found. The best rated items were the ones regarding how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it.

Task 5

		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this general task 5?	Pre	18	2.39	0.698	1	3	0.763	-0.302
	Post	19	2.47	0.697	1	3		
1.2. The user is able to complete the general task 5?	Pre	18	1.56	0.616	0	2	0.480	-0.707
	Post	19	1.42	0.507	1	2		
1.3. Has the user completed the activities of the general task 5?	Pre	16	2.63	0.619	1	3	0.374	-0.889
	Post	19	2.37	0.597	1	3		
2.1. Does the user feel comfortable using the system during the general task 5?	Pre	19	2.58	0.769	0	3	0.739	-0.333
	Post	19	2.58	0.507	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	19	2.42	0.961	0	3	0.084	-1.725
	Post	19	2.79	0.419	2	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	19	2.58	0.902	0	3	0.102	-1.633
	Post	19	2.84	0.375	2	3		
3.1. Do you like the idea of this general task 5?	Pre	17	2.06	0.748	1	3	0.773	-0.289
	Post	15	2.13	0.990	1	3		
3.2. Do you like the general task 5 idea if you future situation changes?	Pre	19	2.32	0.820	1	3	0.206	-1.265
	Post	18	2.17	0.985	0	3		
3.3. Will you buy this concrete system option if it was in the market?	Pre	19	1.16	1.119	0	3	0.256	-1.136
	Post	18	0.89	1.023	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	19	1.84	1.068	0	3	0.782	-0.277
	Post	18	1.83	1.043	0	3		
3.5. Does the user like better other communication device?	Pre	13	0.23	0.439	0	1	0.157	-1.414
	Post	10	0.50	0.527	0	1		

Table 4: Results of the 2nd scenario, task 5
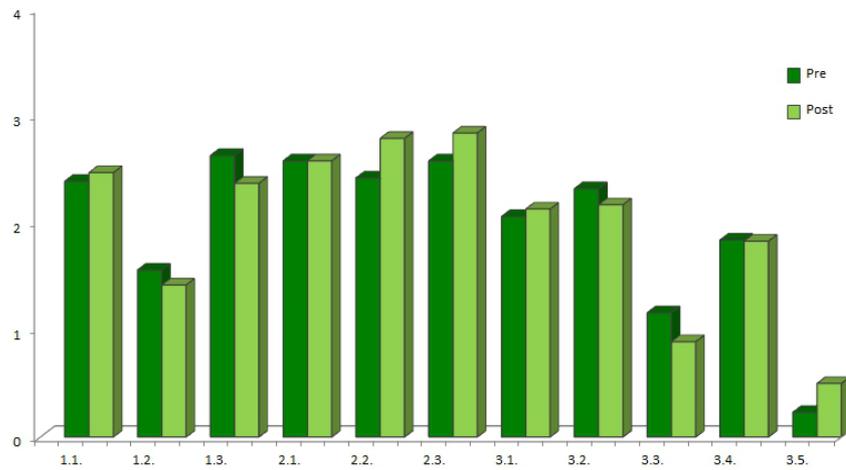
Figure 20: Results of the 2nd scenario, task 5

Even if, in general, scores from the second trial have been higher than the scores of the first trial, there are no significant differences between the scores of both trials. The best rated items were the ones regarding the easiness to complete the tasks, how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it.

Task 6

		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this general task 6?	Pre	18	2.39	0.698	1	3	1	0.000
	Post	18	2.44	0.705	1	3		
1.2. The user is able to complete the general task 6?	Pre	19	1.53	0.513	1	2	0.414	-0.816
	Post	18	1.39	0.502	1	2		
1.3. Has the user completed the activities of the general task 6?	Pre	19	2.37	0.684	1	3	0.366	-0.905
	Post	18	2.22	0.548	1	3		
2.1. Does the user feel comfortable using the system during the general task 6?	Pre	19	2.63	0.684	1	3	0.317	-1.000
	Post	16	2.50	0.516	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	19	2.74	0.653	1	3	0.317	-1.000
	Post	16	2.75	0.577	1	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	19	2.79	0.535	1	3	0.180	-1.342
	Post	16	2.94	0.250	2	3		
3.1. Do you like the idea of this general task 6?	Pre	17	2.24	0.752	1	3	0.157	-1.414
	Post	13	2.46	0.877	1	3		
3.2. Do you like the general task 6 idea if you future situation changes?	Pre	18	2.39	0.850	1	3	0.417	-0.811
	Post	15	2.13	1.187	0	3		
3.3. Will you buy this concrete system option if it was in the market?	Pre	19	1.26	1.284	0	3	0.852	-0.187
	Post	15	1.33	1.175	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	19	1.95	1.177	0	3	0.433	-0.784
	Post	15	2.27	0.961	0	3		
3.5. Does the user like better other communication device?	Pre	11	0.45	0.522	0	1	0.083	-1.732
	Post	8	0.75	0.463	0	1		

**Table 5: Results of the 2nd scenario, task 6**
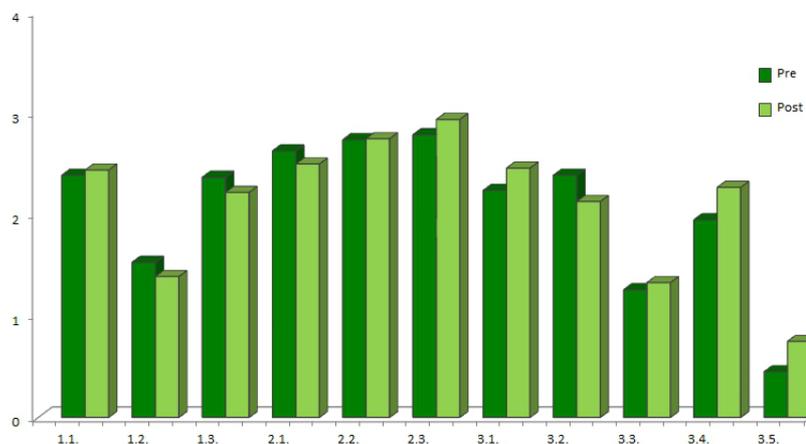
Figure 21: Results of the 2nd scenario, task 6

21

Again, in this case, there are no differences between the first and the second trial. The best rated items were the ones regarding how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it.

Task 7

		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this general task 7?	Pre	17	2.35	0.702	1	3	0.257	-1.134
	Post	18	2.22	0.647	1	3		
1.2. The user is able to complete the general task 7?	Pre	17	1.53	0.624	0	2	0.480	-0.707
	Post	18	1.33	0.485	1	2		
1.3. Has the user completed the activities of the general task 7?	Pre	17	2.29	0.920	0	3	0.763	-0.302
	Post	18	2.17	0.618	1	3		
2.1. Does the user feel comfortable using the system during the general task 7?	Pre	18	2.39	0.916	0	3	0.271	-1.100
	Post	18	2.61	0.502	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	16	2.63	0.719	1	3	0.157	-1.414
	Post	18	2.89	0.323	2	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	16	2.69	0.704	1	3	0.157	-1.414
	Post	18	2.89	0.323	2	3		
3.1. Do you like the idea of this general task 7?	Pre	16	1.88	0.957	0	3	0.084	-1.725
	Post	14	2.36	0.929	1	3		
3.2. Do you like the general task 7 idea if you future situation changes?	Pre	17	1.94	1.029	0	3	0.773	-0.289
	Post	17	1.76	1.200	0	3		
3.3. Will you buy this concrete system option if it was in the market?	Pre	17	0.88	1.111	0	3	0.918	-0.103
	Post	18	0.78	1.114	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	17	1.71	1.105	0	3	0.951	-0.061
	Post	18	1.67	1.188	0	3		
3.5. Does the user like better other communication device?	Pre	7	0.57	0.535	0	1	1	0.000
	Post	8	0.88	0.354	0	1		

**Table 6: Results of the 2nd scenario, task 7**
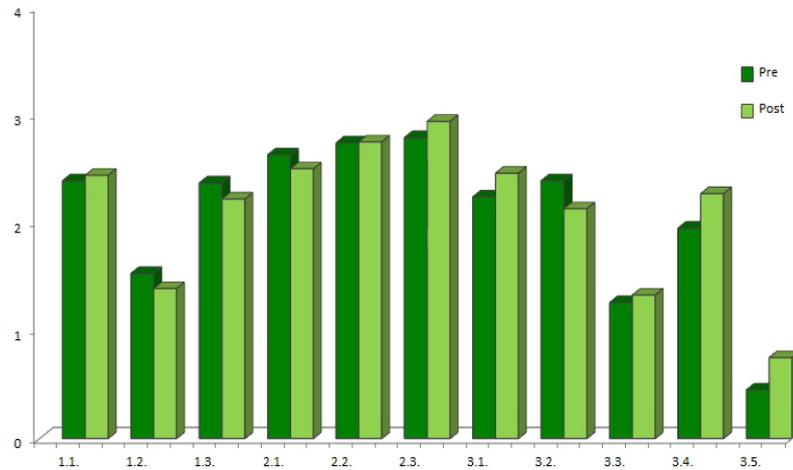
Figure 22: Results of the 2nd scenario, task 7

No significant difference has been found between the two trials in any of the variables. The best rated items were the ones regarding how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it.

Task 8

		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this general task 8?	Pre	13	2.69	0.751	1	3	0.783	-0.276
	Post	19	2.58	0.692	1	3		
1.2. The user is able to complete the general task 8?	Pre	13	1.69	0.480	1	2	0.564	-0.577
	Post	19	1.63	0.496	1	2		
1.3. Has the user completed the activities of the general task 8?	Pre	12	2.58	0.515	2	3	0.705	-0.378
	Post	19	2.47	0.612	1	3		
2.1. Does the user feel comfortable using the system during the general task 8?	Pre	13	2.62	0.870	0	3	0.739	-0.333
	Post	18	2.72	0.461	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	13	2.23	0.927	1	3	0.034	-2.121
	Post	18	2.72	0.461	2	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	13	2.31	0.947	1	3	0.059	-1.890
	Post	18	2.83	0.383	2	3		
3.1. Do you like the idea of this general task 8?	Pre	13	2.15	1.144	0	3	0.339	-0.957
	Post	16	2.50	1.095	0	3		
3.2. Do you like the general task 8 idea if you future situation changes?	Pre	12	2.17	0.937	1	3	0.680	-0.412
	Post	17	1.94	1.144	0	3		
3.3. Will you buy this concrete system option if it was in the market?	Pre	13	1.00	1.291	0	3	0.774	-0.288
	Post	17	1.00	0.935	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	12	1.92	1.240	0	3	0.518	-0.647
	Post	16	1.81	1.223	0	3		
3.5. Does the user like better other communication device?	Pre	6	0.67	0.516	0	1		
	Post	6	1.00	0.000	1	1		
1. Overall. I am satisfied with how easy of completing the tasks in this scenario	Pre	19	3.11	1.629	1	6	0.308	-1.019
	Post	17	2.65	1.455	1	6		
2. Overall. I am satisfied with the amount of time it took to complete the tasks in this scenario	Pre	19	3.63	1.499	1	6	0.751	-0.318
	Post	17	3.53	1.505	1	6		
3. Overall. I am satisfied with the support information (online-line help, Messages, documentation, and training material when completing the tasks.	Pre	19	2.95	1.545	1	6	0.473	-0.717
	Post	17	2.47	1.328	1	5		

Table 7: Results of the 2nd scenario, task 8
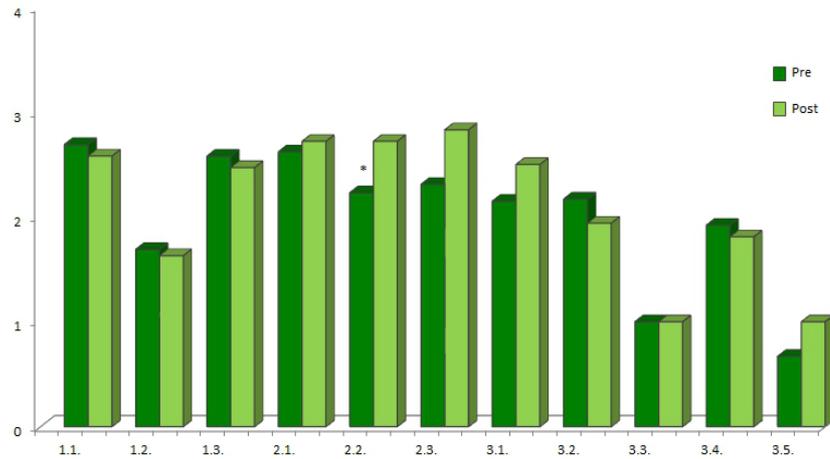
Figure 23: Results of the 2nd scenario, task 8
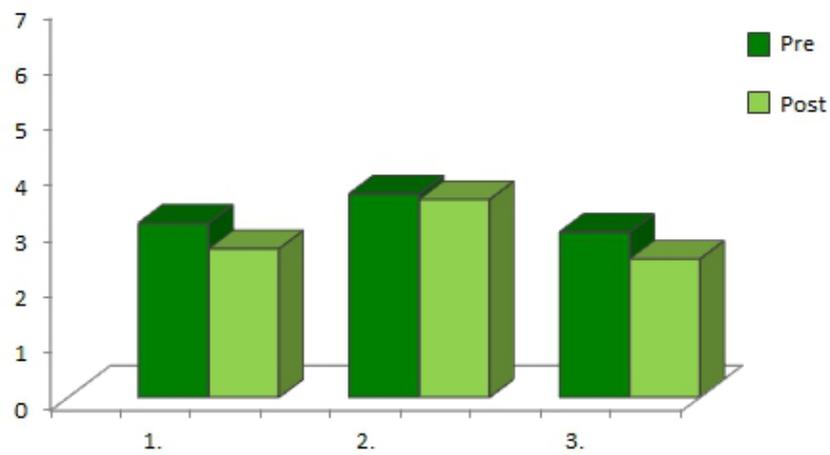
Figure 24: ASQ Scenario 2

In this task, users reported a significant improvement (Z-2.121, $p<0.05$) regarding the differentiation of the colors of the PeerAssist system screen. No difference has been found regarding ASQ results. The best rated items were the ones regarding how easy was to complete the tasks, how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it.

2.3.2.2. Greek results for scenario 2

The Greek results for scenario 2 were similar to the Spanish trials, with a slight overall improvement, justified mainly by the enhanced user interface that provided a more friendly and pleasant way to interact with the system. The best rated items were the ones regarding user friendliness to complete the scenarios, how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it once again similarly to the Spanish trials. To keep the number of graphs reasonably low, we omit detailed results per task and provide only the overall ASQ results below.

1.Usability / 2.Time / 3.Support
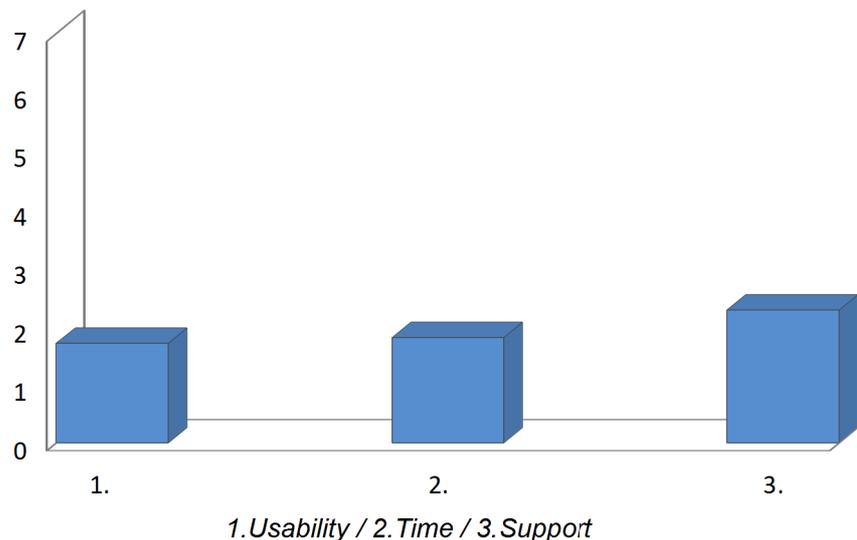
Figure 25: ASQ results of the scenario 2

Likert test (1-7 scale)

2.3.3. Scenario 3: Search and ask for services

Scenario 3 consisted of searching and requesting services with the use of the PeerAssist system. The system offered different kinds of services such as requesting a taxi, hair salon, meals on wheels, etc. The most popular were the taxi service and meals on wheels. In the case of the taxi, once they requested this service, the screen of the PeerAssist system showed the telephone number they had to call to. In the case of meals on wheels, they had to order a meal, e.g. pizza,

and add the specifications they wanted (double cheese, no tuna...).

2.3.3.1. Spanish results for scenario 3

		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this scenario 3?	Pre	18	2.50	0.618	1	3	0.414	-0.816
	Post	19	2.42	0.692	1	3		
1.2. The user is able to complete the general scenario 3?	Pre	18	1.44	0.511	1	2	0.655	-0.447
	Post	19	1.47	0.513	1	2		
1.3. Has the user completed the two activities of the scenario 3?	Pre	18	2.50	0.618	1	3	0.655	-0.447
	Post	18	2.33	0.594	1	3		
2.1. Does the user feel comfortable using the system during the scenario 3?	Pre	19	2.58	0.507	2	3	0.317	-1.000
	Post	19	2.68	0.478	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	19	2.42	0.902	0	3	0.131	-1.511
	Post	19	2.74	0.452	2	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	19	2.47	0.905	0	3	0.131	-1.511
	Post	19	2.79	0.419	2	3		
3.1. Do you like the idea of this concrete scenario 3?	Pre	19	2.21	0.787	1	3	0.480	-0.707
	Post	18	2.11	0.963	1	3		
3.2. Do you like the concrete scenario 3 idea if you future situation changes?	Pre	18	2.11	0.963	0	3	0.874	-0.159
	Post	18	2.11	1.079	0	3		
3.3. Will you buy this concrete system option if it was in the market?	Pre	18	1.06	1.162	0	3	0.952	-0.060
	Post	18	1.06	1.110	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	18	1.78	1.263	0	3	0.327	-0.979
	Post	18	2.06	1.110	0	3		
3.5. Does the user like better other communication device?	Pre	12	0.67	0.492	0	1	0.317	-1.000
	Post	11	0.55	0.522	0	1		
1. Overall. I am satisfied with how easy of completing the tasks in this scenario	Pre	17	2.12	1.269	1	6	0.510	-0.659
	Post	17	2.53	1.419	1	6		
2. Overall. I am satisfied with the amount of time it took to complete the tasks in this scenario	Pre	17	2.82	1.131	1	5	0.721	-0.357
	Post	17	2.76	1.437	1	6		
3. Overall. I am satisfied with the support information (online-line help, messages, documentation, and training material when completing the tasks.	Pre	17	3.00	1.541	1	7	0.905	-0.119
	Post	17	2.88	1.495	1	5		

**Table 8: Results of the 3rd scenario**
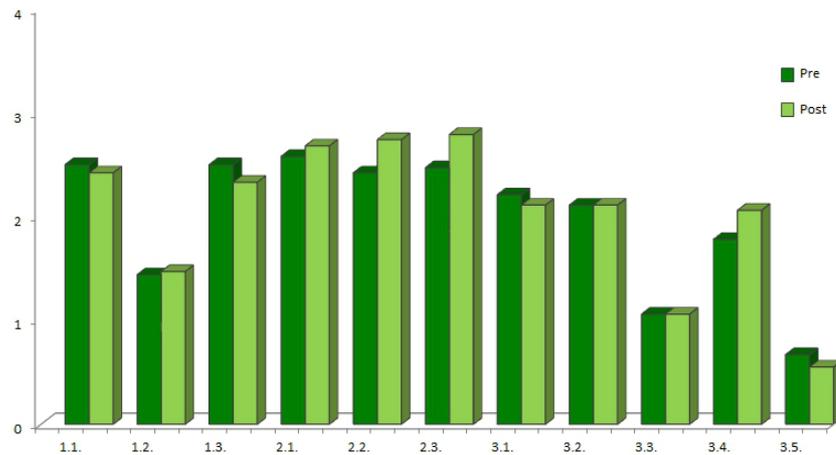
Figure 26: Results of the 3rd scenario
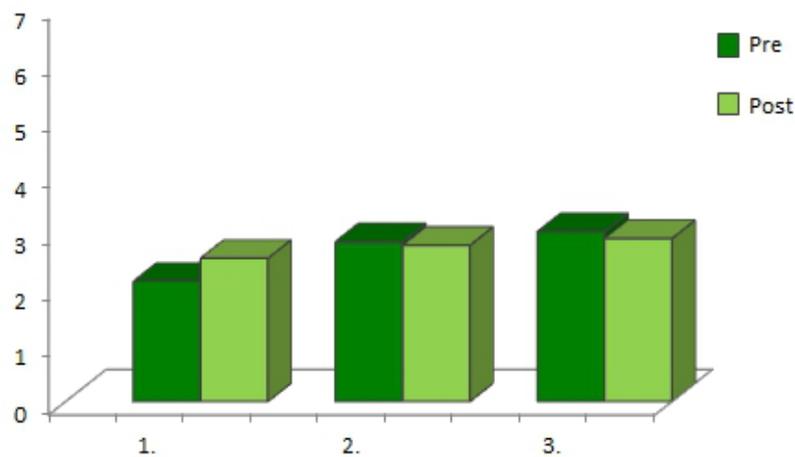
Figure 27: ASQ Scenario 3

No significant difference has been found between the two trials neither in the items of the first questionnaire nor in the second one. The best rated items were the ones regarding how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it. They would like to use it but wouldn't pay for it.

2.3.3.2. Greek results for scenario 3

Again the Greek results were a bit improved compared to the Spanish trials due to the fact that the sample had more computer experience and were additional improvements to the user interface. The best rated items were the ones regarding how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it. They would like to use it but wouldn't pay for it.
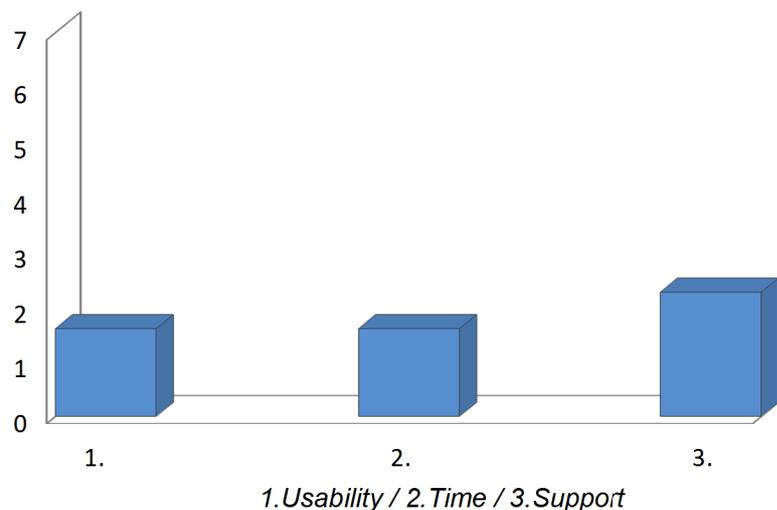
Figure 28: ASQ results of the 3rd scenario

Likert test (1-7 scale)

2.3.4. Scenario 4: Caregivers

The scenario 4 could not be evaluated in Spain, because of some technical problems identified, that were corrected before the Greek trials. The Greek tests in scenario 4 where performed with a assistant taking the role of a caregiver and executing the relevant scenarios together with the trial users. The results were satisfactory and the users liked the idea that they could be able to have someone in contact in case of emergency of just to discuss concerns or everyday problems. What they suggested as a useful application in this scenario, is to be able to have regular sessions through PeerAssist with doctors from a remote location on medical issues. In Athens, this could be implemented provided the system is installed to friendship clubs and a collaboration is initiated with a university hospital to organize the sessions. In face, this is a direction of exploitation we are investigating at the moment.

29
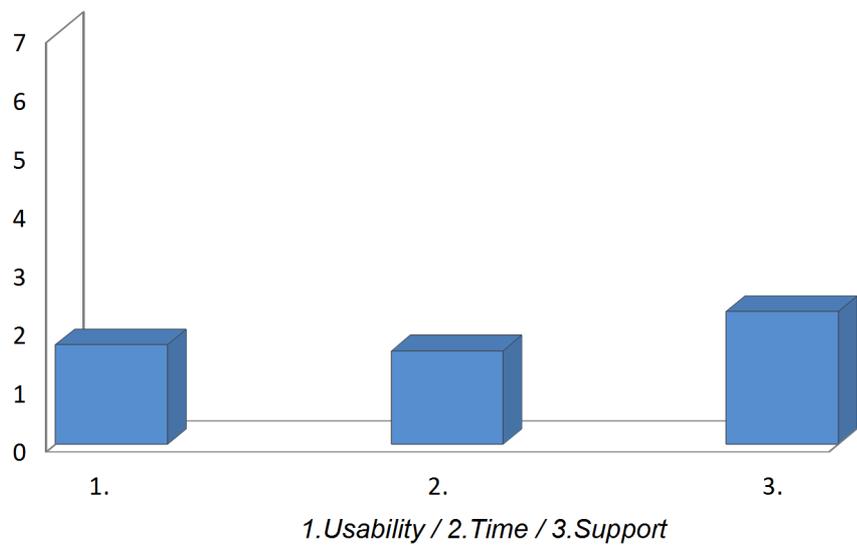
Figure 29: ASQ results of the 4th scenario

Likert test (1-7 scale)

2.3.5. Scenario 5: Notifications

In this scenario the task consisted of checking the different types of notifications (requesting to take part in a chat, a group invitation, friendship invitation, etc.) in the screen of the PeerAssist system.

2.3.5.1. Spanish results for scenario 5

|		N	Mean	Std. Deviation	Min	Max	p	Z
1.1. Has the user been able to interact with the PeerAssist system in this general task 12?	Pre	19	2.53	0.772	1	3	1	0.000
	Post	18	2.50	0.707	1	3		
1.2. The user is able to complete the general task 12?	Pre	19	1.63	0.496	1	2	0.655	-0.447
	Post	18	1.67	0.485	1	2		
1.3. Has the user completed the activities of the general task 12?	Pre	19	2.63	0.684	1	3	0.414	-0.816
	Post	18	2.50	0.707	1	3		
2.1. Does the user feel comfortable using the system during the general task 12?	Pre	18	2.67	0.686	1	3	0.317	-1.000
	Post	18	2.78	0.428	2	3		
2.2. Has the user experimented problems to differentiate the colors of the interface?	Pre	19	2.53	0.964	0	3	0.483	-0.702
	Post	18	2.39	0.778	1	3		
2.3. Has the user experimented problems to differentiate the fonts and texts?	Pre	19	2.05	1.311	0	3	0.040	-2.058
	Post	18	2.67	0.594	1	3		
3.1. Do you like the idea of this general task 12?	Pre	14	1.36	0.842	0	2	0.305	-1.027
	Post	11	1.73	1.104	0	3		
3.2. Do you like the general task 12 idea if you future situation changes?	Pre	18	1.56	1.294	0	3	0.792	-0.264
	Post	16	1.56	1.365	0	3		
3.3. Will you buy this concrete system option if it was in the market?	Pre	18	0.83	1.098	0	3	0.831	-0.213
	Post	16	0.81	1.109	0	3		
3.4. Would you take it at home and use as it is at the moment?	Pre	18	1.50	1.339	0	3	0.666	-0.413
	Post	16	1.44	1.365	0	3		
3.5. Does the user like better other communication device?	Pre	6	0.67	0.516	0	1	0.157	-1.414
	Post	6	0.50	0.548	0	1		
1. Overall. I am satisfied with how easy of completing the tasks in this scenario	Pre	17	2.41	1.583	1	6	0.748	-0.321
	Post	16	2.69	1.493	1	5		
2. Overall. I am satisfied with the amount of time it took to complete the tasks in this scenario	Pre	17	2.59	1.543	1	6	0.927	-0.092
	Post	16	2.56	1.315	1	5		
3. Overall. I am satisfied with the support information (online-line help, messages, documentation, and training material when completing the tasks.	Pre	17	3.41	2.063	1	7	0.178	-1.348
	Post	16	2.50	1.265	1	5		

Table 9: Results of the 5th scenario
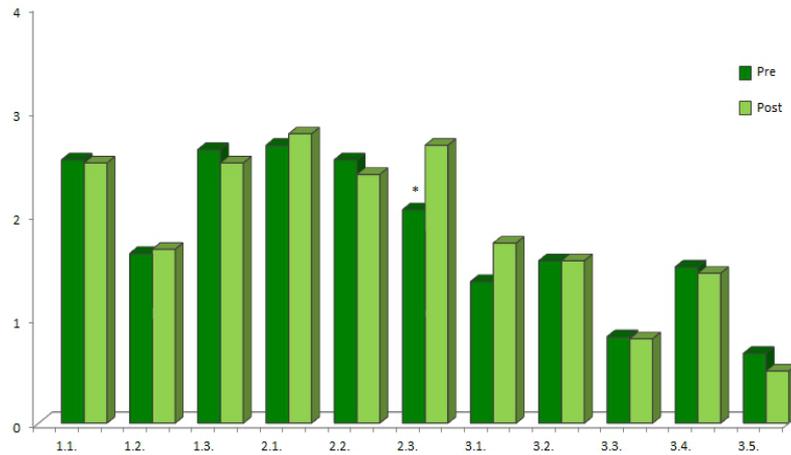
Figure 30: Results of the 5th scenario
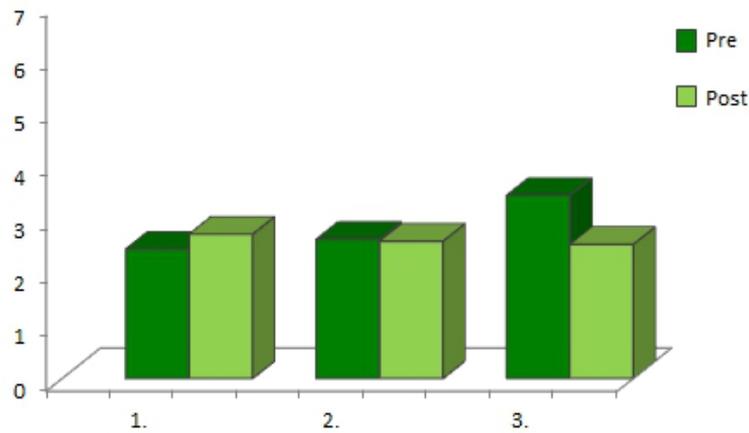
Figure 31: ASQ Scenario 5

Significant improvement (Z=-2.058, p<0.05) has been found between the first and the second trial, after getting familiar with the texts and colors in the notifications. No real difference has been found regarding the ASQ scores. The best rated items were the ones regarding how easy was to complete the tasks, how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it.

2.3.5.2. Greek results for scenario 5

Again the Greek results were better compared to the Spanish trials due to the fact that the sample had more computer experience and additional improvements were made to the user interface, based on the experience from the Spanish trials. The best rated items were the ones regarding how comfortable they feel, how easy was to differentiate the colors on the interface and how easy was to differentiate the fonts and texts. The worst rated was the one regarding the idea of buying it. They would like to use it but wouldn't pay for it, at least at this stage.


1.Usability / 2.Time / 3.Support
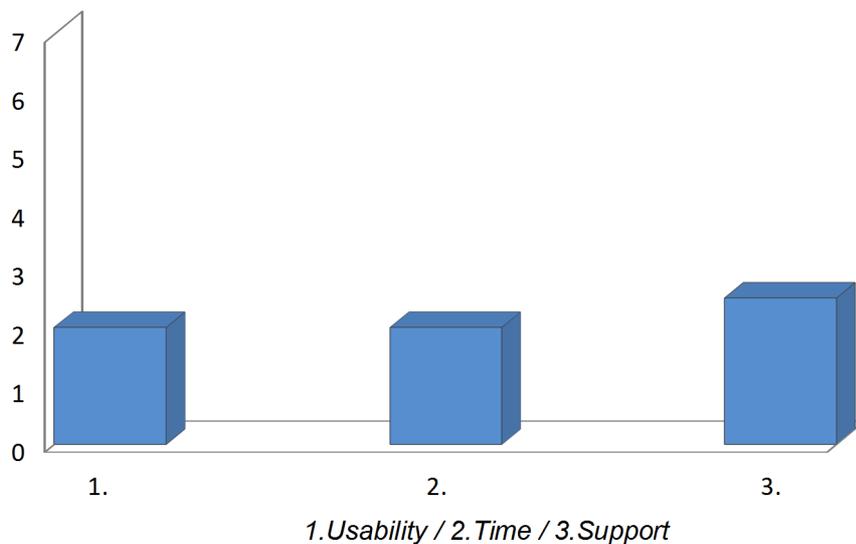
Figure 32: ASQ results of the 5th scenario

Likert test (1-7 scale)

2.3.6. PSSUQ Results

In this section the results obtained in the Post-Study System Usability Questionnaire (PSSUQ) are shown. This 7-point Likert scale questionnaire provides an overall evaluation of the system.

2.3.6.1 PSSUQ Spanish trials

|		N	Mean	Std. Deviation	Min	Max	p	Z
Overall user satisfaction with the system	Pre	19	2.9677	0.82637	1.18	4.47	0.107	-1.610
	Post	19	2.4525	0.87235	1.00	4.17		
System usefulness	Pre	19	3.0987	1.05868	1.00	5.00	0.059	-1.888
	Post	19	2.4474	0.90936	1.00	4.25		
Information quality	Pre	19	2.8678	0.88086	1.00	4.29	0.432	-0.785
	Post	19	2.6575	1.10382	1.00	4.86		
Interface quality	Pre	19	2.8947	0.94315	1.00	4.67	0.021	-2.302
	Post	19	2.1579	1.12419	1.00	5.00		

Table 10: PSSUQ Results
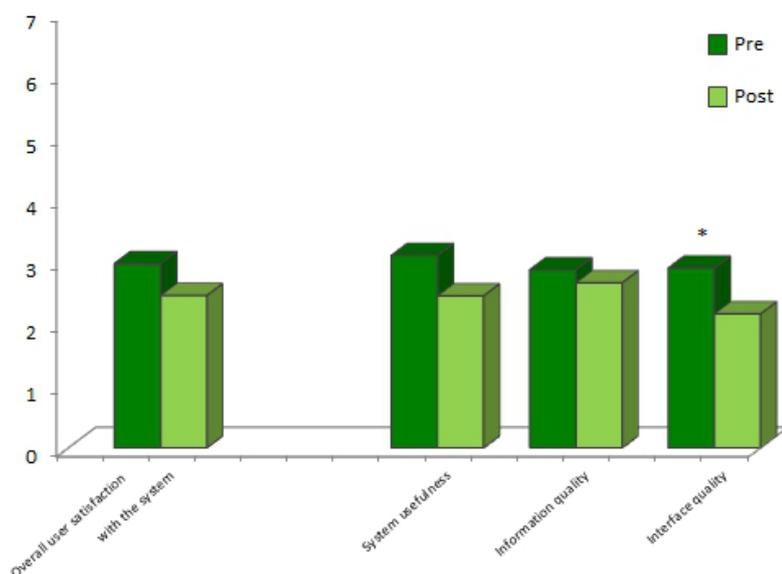
Figure 33: PSSUQ Results – Spanish trials

A significant difference was found on the PSSUQ scores regarding the interface quality (Z=-2.302, p<0.05) being the score of the second trial lower than the score on the first one. Even if the users haven't had any difficulty on distinguishing the text, fonts and colors, they would like to have a better interface quality.

2.3.6.2 PSSUQ Greek trials
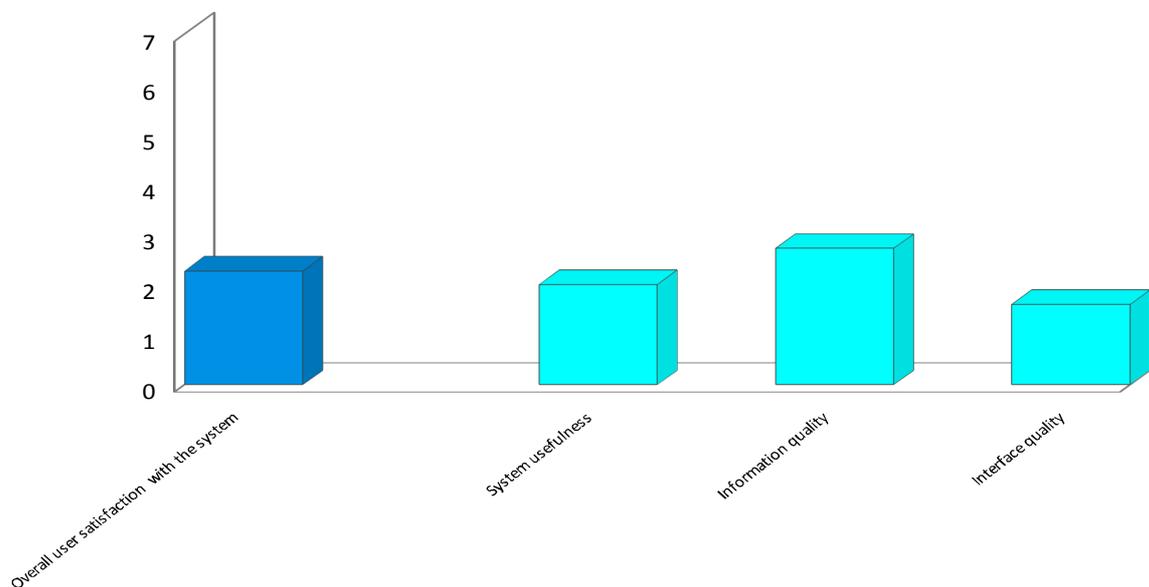
Figure 34: PSSUQ Results – Greek trials

The users did not have any difficulty on distinguishing the text, fonts and colors but they would also like to see an improvement on the user interface quality. Still an improvement to the Spanish trials, but most probably attributed due to the fixes and additions in version used for the Greek trials.

3. CONCLUSIONS

As the users were people from 60 to 82 years old and not all of them very skillful using a computer, the general impression have been lower than if we had been working with more experienced users, but our intention was to have the greater mix of users possible. As an example, it has to be mentioned that many users did not know terms such as blog, chat, video call, etc. Because of that, some of them needed the help of the assistants to carry out the different tasks: complete the profile, choose the options, etc. However, some users carried out the different tasks without much help, including the users that took the PeerAssist terminals at home and run the trials from there.

Many of the users reported that, with short further training and more hands on experience, they could use the PeerAssist system without any difficulty. Less experienced users said that they prefer the phone when they have to communicate with just one person, but if they had to communicate the same message to many people, they would choose the group chat or the video call. Some of these users said that they would rather prefer the video call compared to the chat, because they are not used to write with a keyboard or have difficulties to communicate by writing, and they don't want to expose this difficulty to others. All of the users highly appreciated the emergency notifications to the caregivers as a way to get immediate help in case of need, and also the availability to have access to external services.

The differences between the Greek and Spanish results were primarily due to the improved user interface and the relative better familiarity of the Greek users with the computers. Especially in Spain, less experienced users needed the help of the assistants in order to accomplish the scenarios.

Finally, despite the fact that the users liked the idea of the system, they would hardly pay for buy it at this stage. Based on this fact, exploitation plans are currently targeting other sources of revenue, such as advertisements, external services, public funding, etc.

To conclude, the evaluation outcome was very promising. For a trial version, the users appreciated its usefulness and would really wait to see it in real use. Despite the small time of training, they were able to explore all its capabilities and use it relatively easy. Nevertheless, for large-scale use, special attention should be paid to the training needs of the users, especially non-experienced ones. The weaknesses identified can all be easily covered with moderate effort from the consortium.

REFERENCES

[1] Wilcoxon, Frank (Dec 1945). "Individual comparisons by ranking methods". Biometrics Bulletin 1 (6): 80–83.