AAL Project no: AAL-call-2018-150

VIRTUAAL

Virtual and Augmented reality for combating cognitive impairment

DELIVERABLE nº 2.2

Dashboards and KPIs

Project ref no	AAL-call-2018-150
Project acronym	**VIRTUAAL**
Project full title	**Virtual and Augmented reality for combating cognitive impairment**
Nature[1]	WP2
Dissemination level[2]	Select an option
Due date of deliverable	25/10/2019
Actual submission date	25/10/2019
Deliverable name	Dashboards and KPI Performances
Status	Completed
WP contributing to the deliverable	2
Main contributors	University of Deusto
Other contributors	
Author(s)	Begoña García Zapiraín
Keywords	Test

Abstract (for dissemination)	The role of this deliverable is to present the results of the testing phase and the key performance indicators, which are crucial for the understanding of the overall development of the project.

[1] L Legal agreement, O = Other, P = Plan, PR = Prototype, R = Report, U = User scenario

[2] PU = Public, PP = Restricted to other programme participants (including the Commission Services), RE = Restricted to a group specified by the consortium (including the Commission Services), CO = Confidential, only for members of the consortium (including the Commission Services)

Disclaimer

The information in this document is subject to change without notice. Company or product names mentioned in this document may be trademarks or registered trademarks of their respective companies.

Version History

Version	Edited by	Date	Description
0.1	DEU	14.05.2019	Generation of the document
0.2	DEU	15.05.2019	First version
0.3	DEU	26.05.2019	Second version
0.4	DEU	30.05.2019	Third Version
0.5	DEU	25.10.2019	Fourth Version
1.0	DEU	16.11.2019	Fifth Version
1.1	DEU	11.12.2019	Sixth Version

Table of Content

1. ABOUT THIS DOCUMENT

1.1. Role of the deliverable

This deliverable provides an exhaustive description of both the key performance indicators and dashboards that include the results of the analysis carried out throughout the project. The first chapter describes the objectives of the project and to what extent they have been achieved. Then, the results of the accomplishment of these objectives are depicted in different dashboards and information tables. The role of this deliverable, in sum, is to collect the data and present it in an understandable way.

1.2. Relationship to other VirtuAAL deliverables

This document is closely related to D1.3, where the tests were designed and the basis of the project was established. Once the objectives of this deliverable have been accomplished, these conclusions can be achieved. D2.1 is also crucial for the achievement of these results.

2. Pilots

ANA.

In Romania we organised 16 rounds of trials - with an average session length of 3,5 hours (the shortest session, with one volunteer was of 1 hour, whilst the longest 3 sessions lasted up to 5 hours).

The sessions took place in 2 Senior's Clubs from Bucharest and at our (ANA's) premises - VR training and testing being performed under the direct supervision of our research team, which always included at least one MD, at some sessions 2 MD's (beside the researchers handling the technological aspects and the supportive actions towards the end-users during testings).

No.	Date	Location	No of PEU's	Duration (hours)
1	02-Aug-2019	ANA Premises	3 PEU's (1st session)	3
2	08-Aug-2019	Senior's Club S1	5 PEU's (1st session)	4
3	13-Aug-2019	Senior's Club S1	6 PEU's (3 for 1st session, 3 for 2nd)	4
4	22-Aug-2019	Senior's Club S1	9 PEU's (4 for 1st session, 3 for 2nd, 2 for 3rd)	4
5	29-Aug-2019	Senior's Club S1	9 PEU's (3 for 1st session, 3 for 2nd, 2 for 3rd, 1 for final evaluation)*	4-5
6	05-Sept-2019	Senior's Club S1	5 PEU's (4 for 1st session, 1 for 3rd)	3
7	06-Sept-2019	ANA Premises	6 PEU's (3 for 1st session, 3 for 2nd)	3
8	09-Sept-2019	Senior's Club S6	5 PEU's (1st session)	4

9	12-Sept-2019	Senior's Club S1	12 PEU's (3 for 1st session, 5 for 2nd, 3 for 3rd, 1 for final evaluation)	5
10	16-Sept-2019	Senior's Club S6	5 PEU's (2nd session)	2
11	19-Sept-2019	Senior's Club S1	14 PEU's (1 for 1st session, 3 for 2nd, 5 for 3rd, 5 for final evaluation)	5
12	20-Sept-2019	ANA Premises	7 PEU's (2 for 2nd session, 5 for 3rd)	4
13	23-Sept-2019	Senior's Club S6	5 PEU's (3rd session & final evaluation)	2,5
14	27-Sept-2019	Senior's Club S1	6 PEU's (3 for 2nd session, 3 for 3rd session & final evaluation)	4
15	03-Oct-2019	Senior's Club S1	4 PEU's (1 for 2nd session, 2 for 3rd session & final evaluation and 1 drop-out)	2
16	04-Oct-2019	ANA Premises	1 PEU (3rd session & final evaluation)	1

CAT

In Denmark we organised 26 days for testing with room for either 2 or 4 users per. day. An hour was allocated for each session, but most session took 40-50 minutes including completion of interviews and questionnaires. This allowed for the team to schedule specific time slots for session 1,2 and 3 for each user.

The Final Trial testing ran from the 7th of august to the 20th of September and included a total of 22 users. All tests were done at Vikærgården in Aarhus, which both have short term rooms in which elderly users are offered different rehabilitation opportunities during their stay, as well as rehabilitation opportunities for people living at home. The users were recruited by one of Vikærgårdens occupational therapists.

Each session was supervised by one observer from the project team who oversaw observations, interview with user and therapist and questionnaire. The observer was also able to assist with any minor technical difficulties that the therapist might experience.

No.	Date	Location	No of PEU's	Duration (hours)
1	07-Aug-2019	Vikærgården	2 PEU's (4 for 1st session, 2 cancelled)	4
2	09-Aug-2019	Vikærgården	2 PEU's (2 for 1st session, 2 cancelled)	4
3	12-Aug-2019	Vikærgården	1 PEU (1 for 2nd session, 1 cancelled)	2
4	13-Aug-2019	Vikærgården	2 PEUs (2 for 2nd session)	2
5	14-Aug-2019	Vikærgården	3 PEU's (1 for 1st session, 1 for 2nd session, 1 for 3rd & Fnial, 1 cancelled)	4
6	16-Aug-2019	Vikærgården	2 PEU's (2 for 3rd & Final session, 2 cancelled)	4
7	19-Aug-2019	Vikærgården	2 PEU's (1 for 1st session, 1 for 2nd)	2
8	20-Aug-2019	Vikærgården	2 PEU's (2 for 1set session)	2
9	21-Aug-2019	Vikærgården	4 PEU's (1 for 1st session, 2 for 2nd, 1 for 3rd & Final)	4
10	23-Aug-2019	Vikærgården	4 PEU's (1 for 2nd session, 2 for 3rd & Final)	4
11	26-Aug-2019	Vikærgården	2 PEU's (1 for 1st session, 1 for 2nd)	2
12	27-Aug-2019	Vikærgården	1 PEU's (1 for 1st session, 1 cancelled)	2
13	28-Aug-2019	Vikærgården	2 PEU's (1 for 2nd session, 1 for 3rd session & Final, 2 cancelled)	4
14	30-Aug-2019	Vikærgården	4 PEU's (3 for 1st session, 1 for 3rd& Final)	4
15	02-Sept-2019	Vikærgården	2 PEU's (2 for 2nd session)	2
16	03-Sept-2019	Vikærgården	1 PEU (1 for 1st session, 1 cancelled)	2

17	04-Sept-2019	Vikærgården	3 PEU's (1 for 2ª session, 2 for 3ª & Final, 1 cancelled)	4
18	06-Sept-2019	Vikærgården	4 PEU's (2 for 1ª, 1 for 2ª, 1 for 3ª & Final)	4
19	09-Sept-2019	Vikærgården	2 PEU's (1 for 2ª session, 1 for 3ª session & Final session)	2
20	10-Sept-2019	Vikærgården	2 PEU's (2 for 1ª session)	2
21	11-Sept-2019	Vikærgården	4 PEU's (2 for 1ª session, 2 for 3ª & Final)	4
22	13-Sept-2019	Vikærgården	4 PEU's (2 for 1ª session, 1 for 2ª session 1 for 3ª & Final)	4
23	16-Sept-2019	Vikærgården	1 PEU (1 for 2ª session, 1 cancelled)	2
24	17-Sept-2019	Vikærgården	1 PEU (1 for 3ª & Final session, 1 cancelled)	2
25	18-Sept-2019	Vikærgården	1 PEU (1 for 2ª session, 1 cancelled)	2
26	20-Sept-2019	Vikærgården	2 PEU's (2 for 3ª & Final session)	4

3. Tests

In this analysis several tests have been used: EQ-5D-5L, PTQ-ANA, GDS, Socio-demographic test, MoCA and SUS. Now, they are explained and some templates are attached in order to achieve clarity.

3.1 EQ-5D-5L

This test consists in five questions about user's mobility, self-care, usual activity, pain or discomfort, anxiety or depression and one more question which consist in the percentage of health the user think he has (it is defined as subjective-health; S-H). The first four questions have to be answered with the following scale:

0	no problems/pain/anxious
1	slight problems/pain/anxious

2	moderate problems/pain/anxious
3	severe problems/pain/anxious
4	unable to do the activity/ extreme pain/ extremely anxious

Table 1 - EQ-5D-5L values

The way possible answers are written means that a higher score imply having a worse health. Then, it can be computed 'how much health' users have (it is defined as parametric health). Then, the mean value between parametric-health and subjective-health is what it is consider the health of the user.

Moreover, correlation between the 'quantity' of health and how much score they got in MoCA test can be studied.

3.2. PTQ – ANA

This test consists in questions about the game with given possible answers. To compute them, values from one to five from the first answer in the left to the last one in the right part are given. This showed the global score in every single item taking the mean value from all users in each group. This test was answered two times: one after the first trial and the other one after the last trial. As this was done, differences between first impressions and final ones can be studied. (GSD was carried out in Romania but not in Denmark, so then this measure is only available for that group)

1 Graphics

Brightness of the scene:

□ too dark □ dark □ just right □ bright □ too bright 3D effect:

□ disturbing □ boring □ don't know □ ok □ very cool Models (house, kitchen, bank):

□ ugly □ boring □ ok □ nice □ beautiful

Lighting:

□ unrealistic □ slightly wrong □ ok □ beautiful □ realistic Game graphics overall opinion:

□ ugly □ not so nice □ ok □ nice □ very appealing

2 Sound Volume

Balance between sound effects (eff.) and music (m.):

□ eff. far too loud □ eff. too loud □ well balanced □ m. too loud □ m.far too loud Sound effects:

□ silly □ boring □ ok □ nice □ very nice

3 Input Steering:

□ too easy □ easy □ just right □ hard □ too hard Game targets understandable:

□ not at all □ no □ partially □ mostly □ no problems Game speed:

□ too slow □ slow □ just right □ fast □ too fast Frying the egg:

□ too easy □ easy □ just right □ hard □ too hard

Introducing the forms in the boxes

□ too easy □ easy □ just right □ hard □ too hard Taking the money in the bank:

□ too easy □ easy □ just right □ hard □ too hard

Grabbing objects with HTC:

□ too easy □ easy □ just right □ hard □ too hard What is the best device?

□ Oculus □ HTC □ Augmented Reality What is the best game?

□ Forms □ Kitchen □ Bank

How satisfied are you with the game experience:

- **Forms**: Very dissatisfied □ Dissatisfied □ Satisfied □ Very satisfied □
- **Bank**: Very dissatisfied □ Dissatisfied □ Satisfied □ Very satisfied □
- **Kitchen- Vive HTC:** Very dissatisfied □ Dissatisfied □ Satisfied □ Very satisfied □
- **Kitchen- Oculus Go:** Very dissatisfied □ Dissatisfied □ Satisfied □ Very satisfied □

Very dissatisfied- 1 point
Dissatisfied- 2 points
Satisfied – 3 points
Very satisfied- 4 points

3.3. GDS

Geriatric Depression Scale (GDS) is made up of the following fifteen questions. The user scores one point for each bolded answer. If he/she gets more than five it suggests depression.

1. Are you basically satisfied with your life? yes **no**
2. Have you dropped many of your activities and interests? **yes** no
3. Do you feel that your life is empty? **yes** no
4. Do you often get bored? **yes** no
5. Are you in good spirits most of the time? yes **no**
6. Are you afraid that something bad is going to happen to you? **yes** no
7. Do you feel happy most of the time? yes **no**
8. Do you often feel helpless? **yes** no
9. Do you prefer to stay at home, rather than going out and doing things? **yes** no
10. Do you feel that you have more problems with memory than most? **yes** no
11. Do you think it is wonderful to be alive now? yes **no**
12. Do you feel worthless the way you are now? **yes** no
13. Do you feel full of energy? yes **no**
14. Do you feel that your situation is hopeless? **yes** no
15. Do you think that most people are better off than you are? **yes** no

The possible correlation between depression and (EQ-5L-5L) health users have can also be studied (GSD was carried out in Romania but not in Denmark, so then this measure is only available for that group).

3.4. Socio-demographic

1. Questionnaire with socio-demographic aspects

1. Name:

2. Age:

3. Gender: Female Male

4. Educational level – please select only the highest level achieved

- High-school (1)
- Professional school (2)
- Superior studies (university) (3)
- Post-university studies (4) ● Other (5) – please specify:

5. Living status – choose only one answer

- Alone (1)
- With spouse only (2)
- With spouse and other family members (3)
- With other members of my family (4)
- In a retirement home (5)
- Other (6) – please specify:

In the following questionnaire, we will ask you about your **interaction** with **new technology**. New technology is defined as new technical devices i.e smartphone, Smart TV, computer, mobile phone applications, laptop, tablet, VR system/games

Please indicate the degree to which you agree/disagree with the following statements.	Totally disagree (1)	Disagree (2)	Agree (3)	Totally agree (4)

01 I like testing the functions of new technologies.				
02 When I have a new technical system in front of me, I try it out intensively.				
03 I enjoy spending time becoming acquainted with a new technology.				
04 I think that new technologies could be useful for my everyday life				

Table 2 - Socio-demographic questions

3. General Questions on technology usage

● How often do you use the following services/devices/applications?

	Never (1)	Rarely (2)	Sometimes (3)	Usually (4)	Always (5)
Internet					
Desktop personal computer					
Laptop					
Conventional phone - no Smart phone					

Smart Phone					
Traditional TV					
Smart-TV					
Tablet					
Email					
Social networks, e.g. Facebook, Twitter					
Mobile phone applications					

Table 3 - Technology usage questions

This test helps to know the **ratio** between **men and women**, the **mean educational level** and **living status** and **general interaction with technologies**.

In these two tables the mean interaction with technologies and their usage; moreover, their possible correlation is shown.

Furthermore, as these tests were answered by Danish and Romanian users, it is possible to study the relationship between them and if there is any significant statistical difference in there.

In order to do this, **Mann-Whitney U test** is been used because knowing what distribution is under data is not possible. So with this test it is possible to know if the two groups come from two different populations or if they do not and come from the same population. There are two parameters: number of elements of the first sample (n) and number of elements of the second sample (m). Our statistic is

$$U = min(U_1, U_2)$$

$$U_1 = n \cdot m + \frac{n(n2+1)}{} + R_1$$

$$U_2 = n \cdot m + \frac{m(m2+1)}{} + R_2$$

where R_1 and R_2 are the sum of the positions that are assigned to the ordered values (ordered from the lowest to the highest).

3.5. MoCA

MoCA test is a brief screening tool for mild cognitive impairment. This tool focuses in the following skills: visuospatial/ executive (scored from 0 to 5), naming (from 0 to 3), attention (which is divided into three parts scored from 0 to 2, 1 and 3, respectively), language (which is also divided but into two parts scored from 0 to 2 and 1, each one), abstraction (from 0 to 2), delayed recall (from 0 to 5) and orientation (from 0 to 6). Maximum score is 30 and if the result is 26 or over it is considered as normal. One extra point has to be added if the user which is tested has studied less than twelve years (and if his score is less than 30).

On one hand, it can be studied if there is any correlation between EQ-5L-5D and MoCA in each group. On the other, it can be studied the possible correlation between Interaction with new technologies and MoCA too.

Again, using the U test, the relationship between Danish group and Romanian group can be measured and if there is any significant statistical difference in there.

3.6. SUS

This is the System Usability Scale. It is a tool which measures, as its name implies, the usability of the games in this case. Users have to answer from strongly disagree (1) to strongly agree (5) to the following ten items:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

These ten items can be divided into two groups: **positive orientated** (**PO**) items and **negative orientated** items (**NO**). PO is formed by those items which accomplish that if you give a high score to them it is good for the game. On the other hand, giving high score to an item from NO is negative for the game. In this case, PO items are 1, 3, 5, 7 and 9 and NO are the rest (blue-coloured in the table).

Later, relations between item in each group (PO or NO), after the first trial and after the last trial, are studied. In order to get this, a **Wilcoxon test** is used. This test consist in study if two paired samples are independent from the variable which is studied. In this case, this variable is the fact of getting used to the game. Then, there are two sets of data: the one of scores after first trial and the one of scores after last trial, where each value of the first set has a corresponding one in the second set. With all this pairs, taking the one of the first set and subtracting the one on the second set, it results a difference. After this, ordered absolute values of these differences gives a concrete position to each original difference (the one whose absolute value is associated with that position). Finally the **statistic** is the **minimum of two values: the sum of the positions associated to positive differences and the one associated to negative differences**.

4. KPIs

KPIs mark the objectives and accomplishments of the project. They are a series of quantitative indicators whose achievement means success in the testing phase. The following table shows these KPIs:

Key Performance Indicator	**Achieved/not**
Achieving 40 testers	Yes
General satisfaction level (as measured by SUS):	Yes
Balanced gender distribution in testers	Yes

Creation of 3 usable and entertaining games for virtual reality with positive feedback in the tests	Yes
Completion of the test phase with all the tests passed	Yes

5.Analysis

VirtuAAL is a project whose objective is developing immersive games in virtual and augmented reality in order to fight cognitive impairment in patients at home and in nursing homes. Moreover, it aspires to studying the impact of new technologies in improving cognitive decline.

In this work, data is collected from several tests which can be used to measure the social status of patients, their cognitive state and the possibility of them having depression. Moreover, it is also studied the usability of games and the technology being used with the results that the patients get playing these games. They were tried in two locations: Denmark and Romania.

This paper contains information about this data: the analysis of every test which was answered in each group, the study of possible relationships between both groups and, finally, all conclusions that arise after this.

5.1. Denmark

In Denmark twenty-two people played the games. Tests which were used were Socio-demographic (answered by fourteen people), EQ-5L-5D (answered by twelve people), MoCA (answered by 19 people) and SUS (answered by 20 people)

5.1.1 Socio-demographic test

Women	Men	Total
11	8	19

Table 4 - Gender distribution

	U5	U7	U8	U9	U11	U12	U13	U14	U15	U17	U18	U20	U21	U22
Educational level	3	3	5	0	4	0	2	0	2	5	2	3	5	3
Living status	2	2	1	1	2	1	1	1	1	1	2	3	4	1

Table 5 - Socio-demographic data

	Educational level	Living status
Mean value	2,64	1,64
SD	1,78	0,92
Main value	3	1

Table 6 - Socio-demographic data

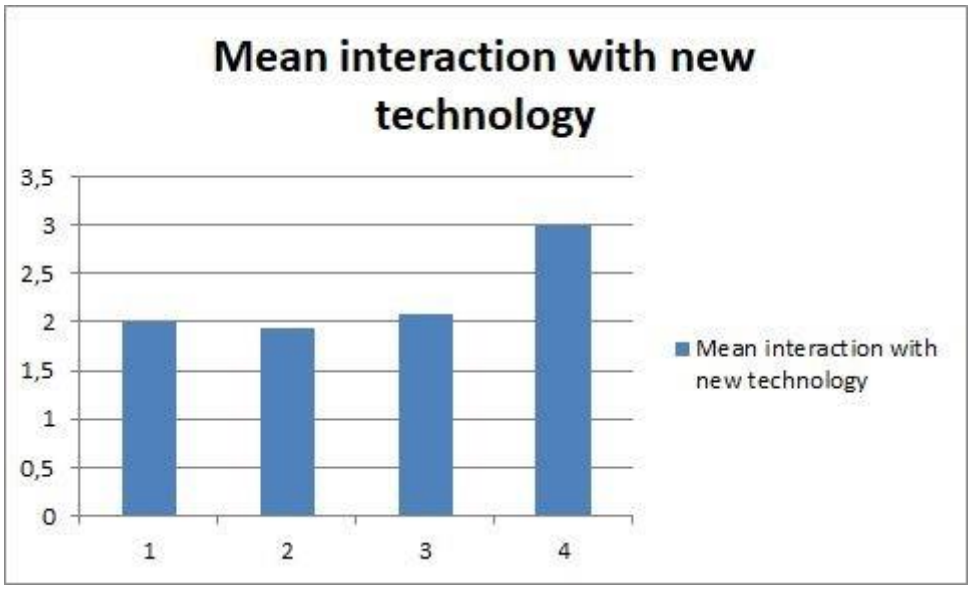


Table 7 - Mean interaction with new technology

Here 1, 2, 3 and 4 are 'I like testing the functions of new technologies.', 'When I have a new technical system in front of me, I try it out intensively.', 'I enjoy spending time becoming acquainted with a new technology.', 'I think that new technologies could be useful for my everyday life.'

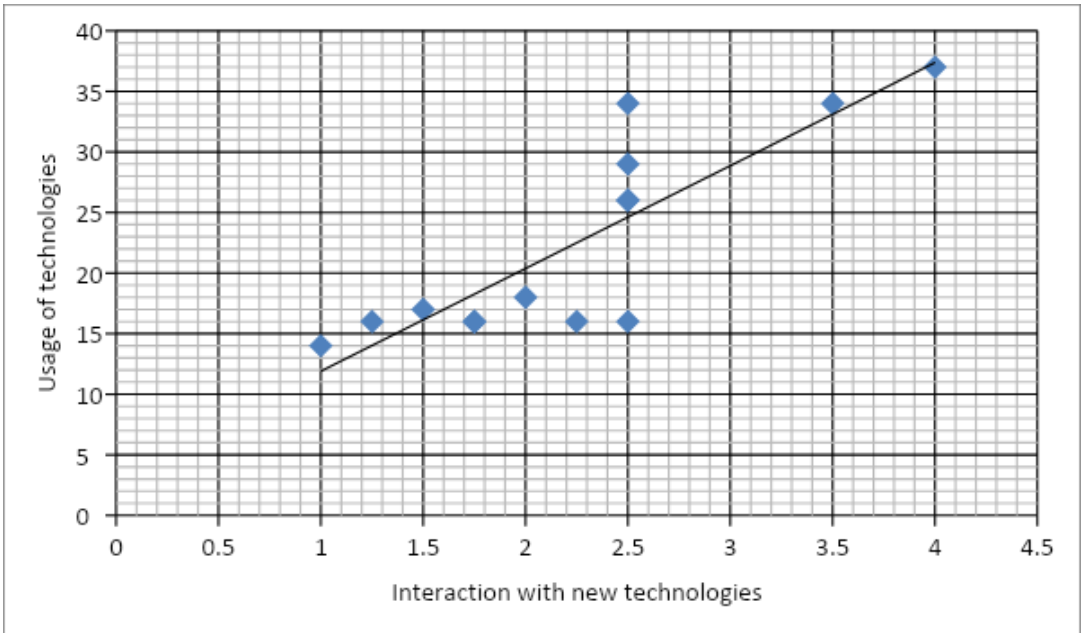


Table 8 - Interaction with new technologies

Only 19 people answered about their gender: 8 were men and 11 were women; this means **57.895%** were **women**. Then, only 14 people answered to educational level and living status question with the following result.

In Denmark, they added the item 'Primary School' (0). As it is seen values go from 0 to 5. The mean education level (2.64 points) is near to superior studies but it cannot be concluded this because standard deviation (SD) is 1.78 points and this means (and it can be proved looking at the data) it moves from 'high-school' to 'other'. This means that it is a multimodal distribution and **there is more than one main educational level**. User number 8, 17 and 21 answered 'tailor (4 years education)', 'similar to nowadays 7th grade but in the old school system in Denmark' and 'have worked on a farm/housework', respectively.

In the part concerning living status, values go from 'living alone' (1) to 'with others members of my family' (4). It is observed that mean level (1.64) is near to the main value (2) and it represent quite good the situation. So then, **the most typical living status is living alone**.

And **for the common use of devices, apps, etc.**, values go from 14 to 37 and it is obtained a mean value of **22.5 points over 55** points they can get in total.

It is very interesting to notice that there is a **high correlation between their interaction with new technologies** (the interaction with new technologies value for a user is defined as the mean of his answers to that question) **and their usage**, as $\rho = 0.8212232$ (Spearman's ρ coefficient).

5.1.2. **EQ**-5L-5D

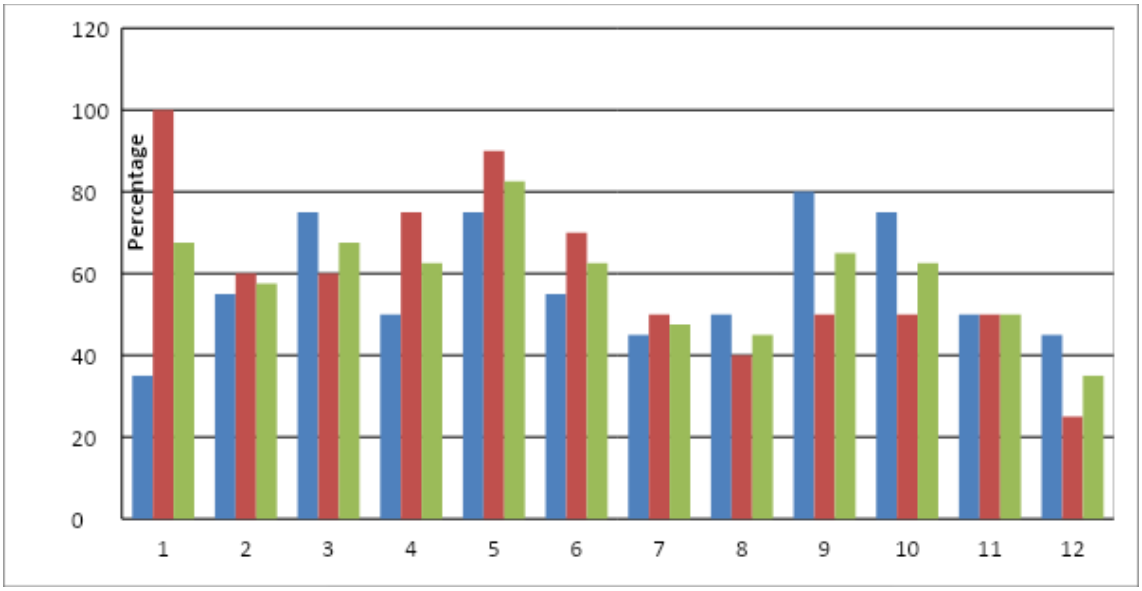


Table 9 - EQ-5L-5D results

This test shows in Denmark a **mean health of 58,75**.

The correlation between EQ-5L-5D and MoCA on one hand and Interaction with new technologies on the other can be computed. The R^2 value is 0,1459 and 0,1193 respectively, so **there is no significant correlation between EQ-5L-5D and MoCA, and between EQ-5L-5D and Interaction with new technologies.**

5.1.3. MoCA

U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U14	U15	U16	U17	U18	U19	U21	U22
23	25	23	24	25	16	28	15	15	25	21	23	17	23	13	18	23	15	22

Table 10 - MoCA results in Danish group

Values go from 13 to 28. It is remarkable that **twelve of them** (they are nineteen) **did not study more than twelve years** (as it is shown in the data). Moreover, as it can be seen, **all of them** (except U7) **are less than 26** points:

eleven are between 21 and 26, six are between 14 and 21 and only U17 is less than 14. Mean value is 20,73 and SD is 4,39, so **they principally are between 16 and 25.**

Just like in the previous case it can be considered the possible correlation between

MoCAis already results done). and In **interaction** this case **R with**$_2$ = **0,37907 new**

technologies which is **not** (the **sufficient** pair MoCA **to claim**and EQ-5L-5D **that any**

correlation exists.

5.1.4. SUS

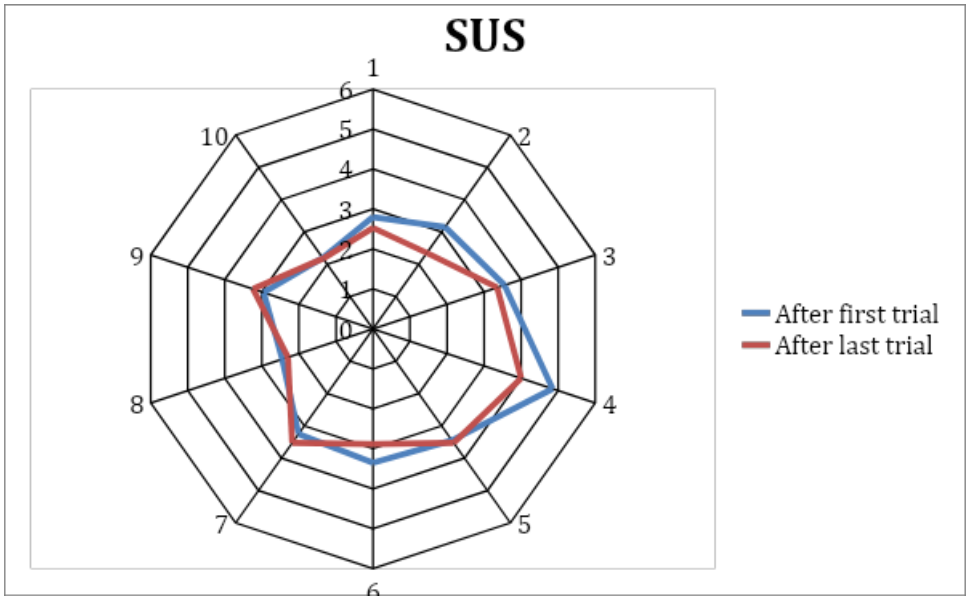


Table 11 - SUS results

	First trial	Last trial
1	2,8	2,52
2	3,15	2,35
3	3,55	3,35

4	4,85	4
5	3,45	3,52
6	3,35	2,88
7	3,25	3,52
8	2,45	2,29
9	2,95	3,23
10	2,2	2,17

Table 12 - Mean scores for each item in SUS tests after first and last trial

It is clear that **items from NO got lower scores**. Using a Wilcoxon, it is seen that with a significance level **α = 0.1 items after first trial and after last trial are statistically different** for items in NO. On the other hand, if a significance level **α = 0.1 is fixed** too, it cannot be asserted that items after first trial and after last trial are statistically different for items in PO.

After this, it can be said that **after getting used with the game it becomes more usable**.

Moreover, almost every item in **PO increased its score**

5.1.5. Kwido

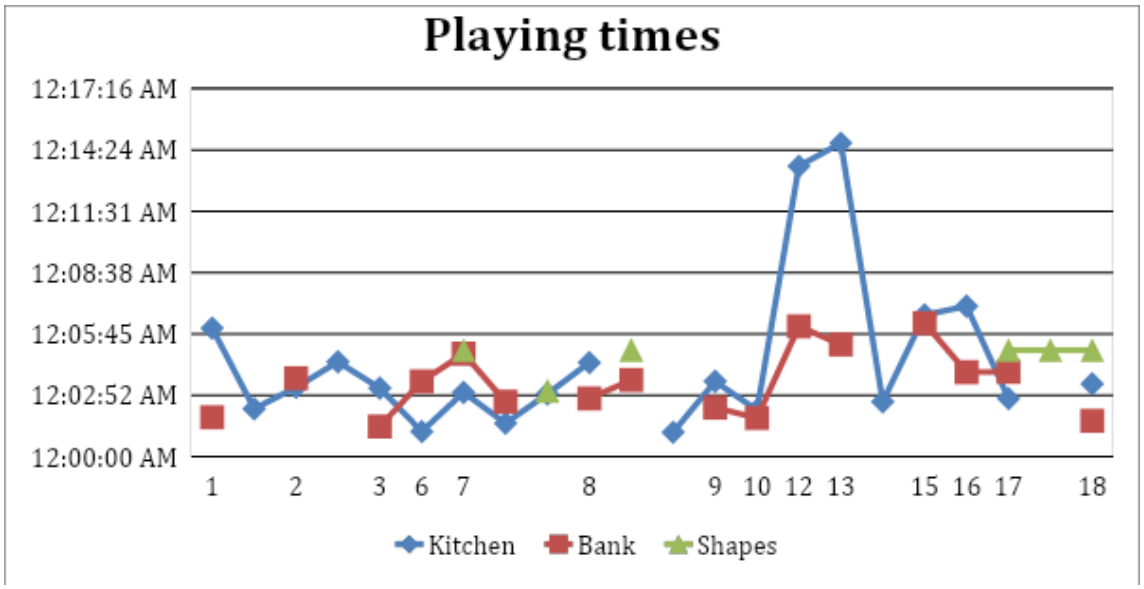


Table 13 - Playing Times

	Kitchen	Bank	Forms
Attempts	21	16	7
Total time	1:42:25	0:55:38	0:33:07
Mean time	0:04:53	0:03:29	0:04:44

Table 14 - Attempts and timing from Kwido platform

U1	U2	U3	U6	U7	U8	U9	U10	U12	U13	U15	U16	U17	U18
73,33	72	72,5	80,5	76	72	82	77,5	80,5	71	62	83,5	59,75	75,8

Table 15 - GCL **(General Cognitive Level) from Kwido platform**

From Kwido platform it is obtained the graphic which shows how many times users played the games and how much it took them. Furthermore, it is seen that the **most played game** was **Kitchen**.

5.1.6. Gender comparison

	n1	n2	U	U(n1, n2, α = 0.05)
Interaction with new technologies	5	6	5	3
Usage of technologies	7	8	24	10
MoCA	9	12	37	19
EQ-5D-5L	4	5	9,5	1

Table 15 - U tests for gender comparison in Denmark

5.1.7. Correlation between MoCA and usability of games

It can be considered if there is any difference between results depending on the gender. Four U test are done in order to know this. As it is shown, **it can not be concluded statistical differences to exist at any relevant significance level for any item**.

	n1 (MoCA under 25)	n2 (MoCA equal or over 25)	U	U(n1,n2,0.05)
SUS	6	14	28	21

Table 16 - U tests for tests comparison between patients under 25 in MoCA test and patients equal or over 25 for MoCA and SUS correlation.

	Entire group	MoCA score under 25
SUS	0.4631325	0.02120194

Table 17 - Spearman's ρ coefficient

In this case, first column represents Spearman's ρ coefficient for the entire group. On the other hand, second column represents Spearman's ρ coefficient for patients whose MoCA score is under 25 points.

It can not be concluded statistical differences to exist between both groups in SUS test.

Moreover, there are no significant correlation between variables.

5.2. Romania

In Romania thirty-three people played the games. Tests which were used were
Socio-demographic), EQ-5L-5D, MoCA, SUS (one for Oculus and other for HTC
Vive), PTQ-ANA

5.2.1. Socio-demographic

Women	Men	Total
25	8	33

Table 18 - Gender distribution

	Education level	Living status
Mean value	2,66	1,73
SD	1,16	0,67
Main value	3	2

Table 19 - Socio demographic data

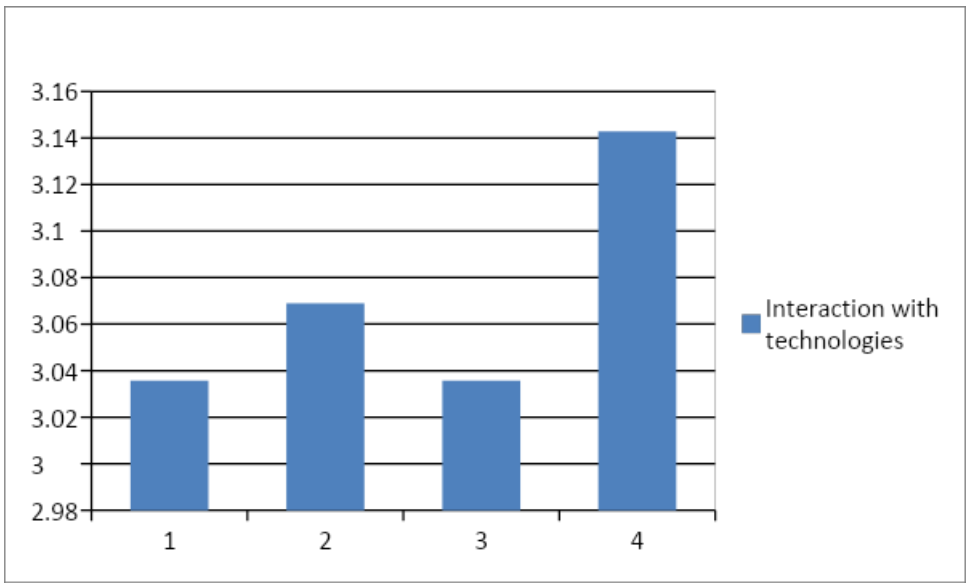


Table 20 - Interaction with technologies

Here 1, 2, 3 and 4 are again 'I like testing the functions of new technologies.', 'When I have a new technical system in front of me, I try it out intensively.', 'I enjoy spending time becoming acquainted with a new technology.', 'I think that new technologies could be useful for my everyday life.'

25 Of the total were women and 8 were men; this means **75,75%** were **women**. All of them answered to education level and living status questions with the following result. Educational level values go from 'high-school' (1) to 'other' (5). The mean education level (2,66 points) is near to 'superior studies' (3), standard deviation is 1,16 points and this mean it can go from 'professional school' studies to 'post university' studies. But in this case, the main value is '**superior studies**' (3) which appears sixteen times, so it can be said that this is the **most common educational level**. Only user number 29 answered 'other' and more precisely '8 years of schooling'.

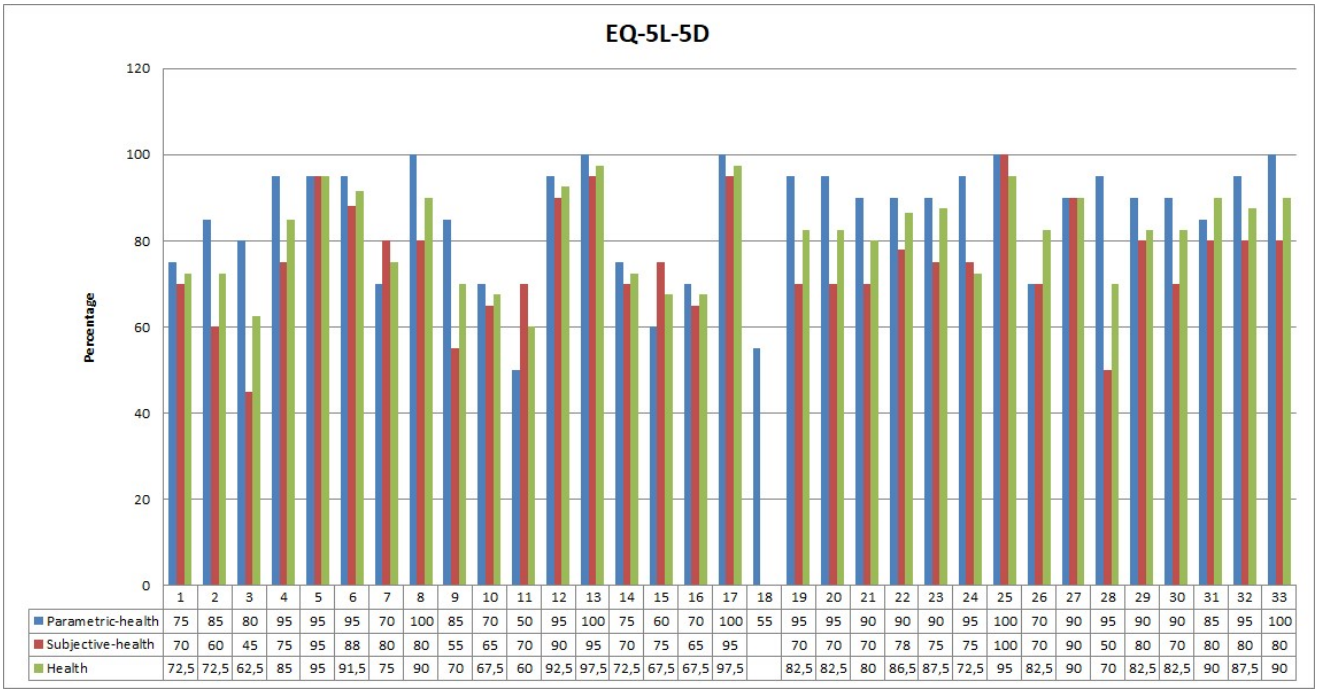
This time, studying living status, whose values go from (1) to 'spouse and family' (3), it is got a mean level of 1,73, a standard deviation of 0,67 points and a main value which is **living with spouse** (2). It appears sixteen times too, so it can be asserted that this is the **most common living status**.

All mean interaction with new technologies are over 3, so it is a **relatively high interaction with new technologies level**.

The **common usage** of devices, apps, etc., moves from 8 to 41, but 8 only appears one time and the next higher value is 16. It is obtained a mean value of **26,06 points over 55** points they can get in total.

In this case, it cannot be asserted that there is any correlation; even more, they are in fact **almost independent**. (**ρ = 0.2422732**)

5.2.2. EQ-5L-5D



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Parametric-health	75	85	80	95	95	95	70	100	85	70	50	95	100	75	60	70	100	55	95	95	90	90	90	95	100	70	90	95	90	90	85	95	100
Subjective-health	70	60	45	75	95	88	80	80	55	65	70	90	95	70	75	65	95		70	70	70	78	75	75	100	70	90	50	80	70	80	80	80
Health	72,5	72,5	62,5	85	95	91,5	75	90	70	67,5	60	92,5	97,5	72,5	67,5	67,5	97,5		82,5	82,5	80	86,5	87,5	72,5	95	82,5	90	70	82,5	82,5	90	87,5	90

Table 21 - EQ-5L-5D results

This test shows in Romania that health oscillates between 62,5 and 97,5. The **mean health** is **80,9**.

As it been said it can computed the correlation between EQ-5L-5D and MoCA on one hand and **Interaction with new technologies on the other**:

EQ-5L-5D and **MoCA**: **ρ = 0.3321319**.

EQ-5L-5D and **Interaction with new technologies on the other:**
ρ = -0.06436213.

Then it is **far from exist significant correlation**.

5.2.3. MoCA

U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U17
24	26	22	29	27	28	27	26	28	25	25	30	28	26	30	24

U18	U19	U20	U21	U22	U23	U24	U25	U26	U27	U28	U29	U30	U31	U33	U34	U35
26	23	26	23	22	30	26	28	28	25	29	26	27	28	20	24	28

Table 22 - MoCA results in the Romanian group

Values moves **from 20 to 30**. It is extraordinary that **any of them is under 20**: twentytwo are greater or equal to 26 and eleven are between 20 and 26. Mean value is **26,1818** and SD is **2,4806**, so **they principally are between 24 and 28.**

Just like in the previous case it can be considered the possible **correlation** between **MoCA** results and **interaction with new technologies**.

MoCA results and **interaction with new technologies:** ρ = **-0.1039377** This is **not** either **sufficient** to claim that any correlation exists.

5.2.4. GSD

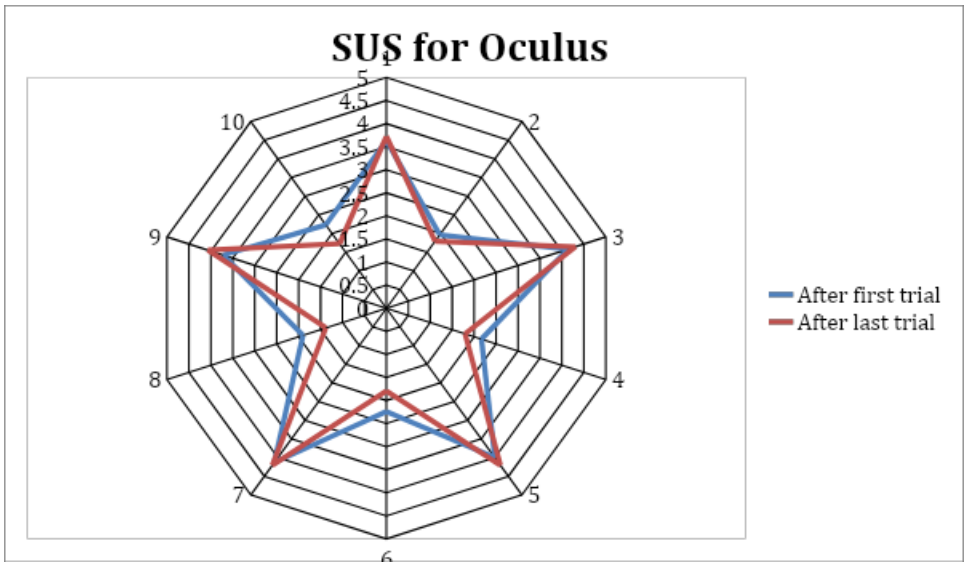
U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U17
5	5	1	6	0	6	5	2	4	4	1	1	0	4	2	2

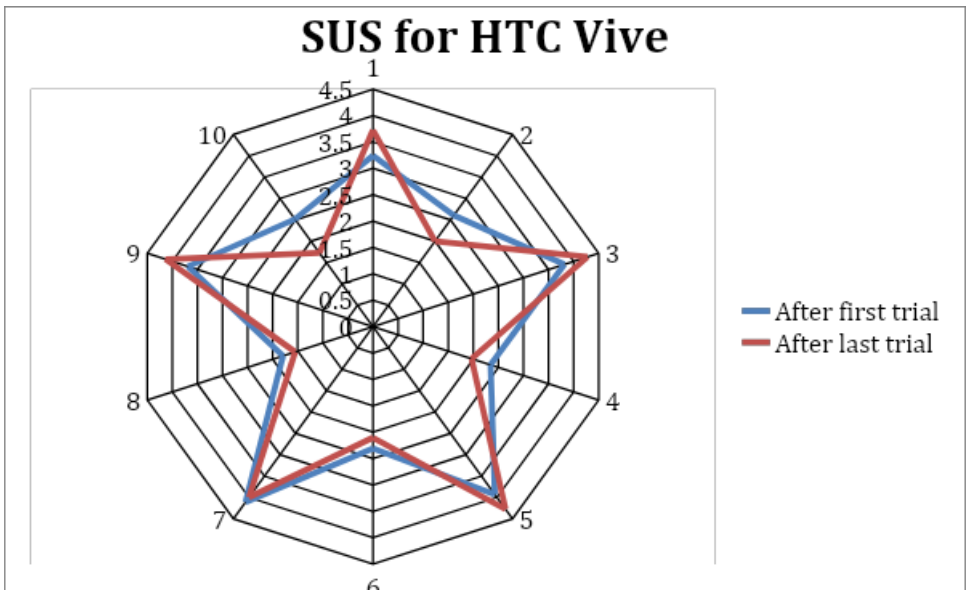
U18	U19	U20	U21	U22	U23	U24	U25	U26	U27	U28	U29	U30	U31	U33	U34	U35
1	8	1	4	7	6	0	7	0	4	0	9	6	1	1	8	1

Table 23 - GDS results in Romanian group

This table shows that there are **nine people** (27,2727 %) who **could have depression**.

5.2.5. SUS





Tables 24 & 25 - SUS for HTC and Oculus

	Oculus		HTC Vive	
	First trial	Last trial	First trial	Last trial
1	3,63	3,72	3,24	3,72
2	1,96	1,79	2,607	2
3	4,13	4,31	3,81	4,2
4	2,16	1,79	2,34	1,96
5	4	4,206	3,93	4,27
6	2,24	1,79	2,31	2,10
7	4,13	4,206	4,11	4
8	1,9	1,37	1,79	1,55
9	3,72	4,06	3,68	4,13
10	2,23	1,72	2,51	1,72

Table 26 - Technology usage questions, mean scores for each item in SUS tests for Oculus and HTC Vive after first and last trial

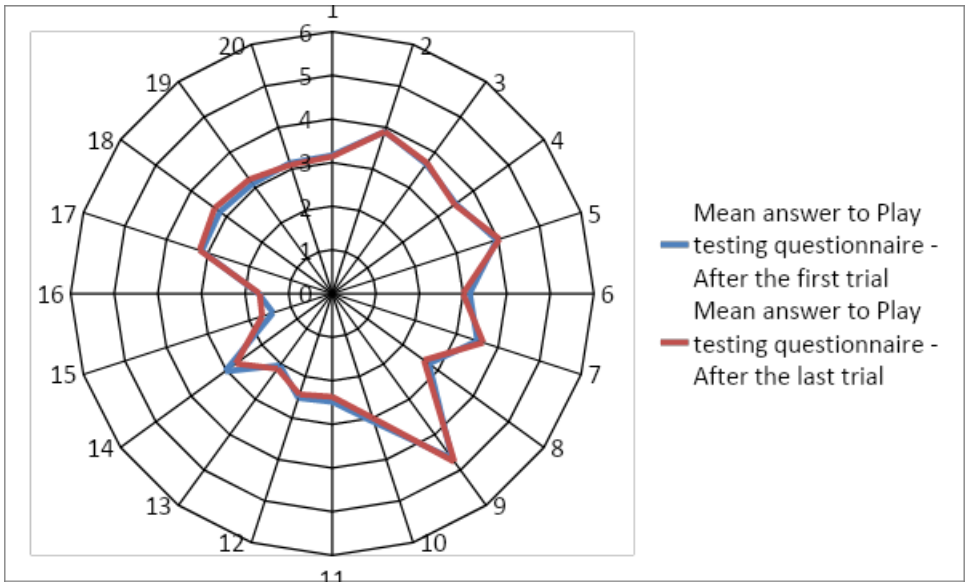
As it is noticed before, the Romanian partner did one SUS test for Oculus and other for HTC Vive. Again, the defined sets PO and NO are considered.

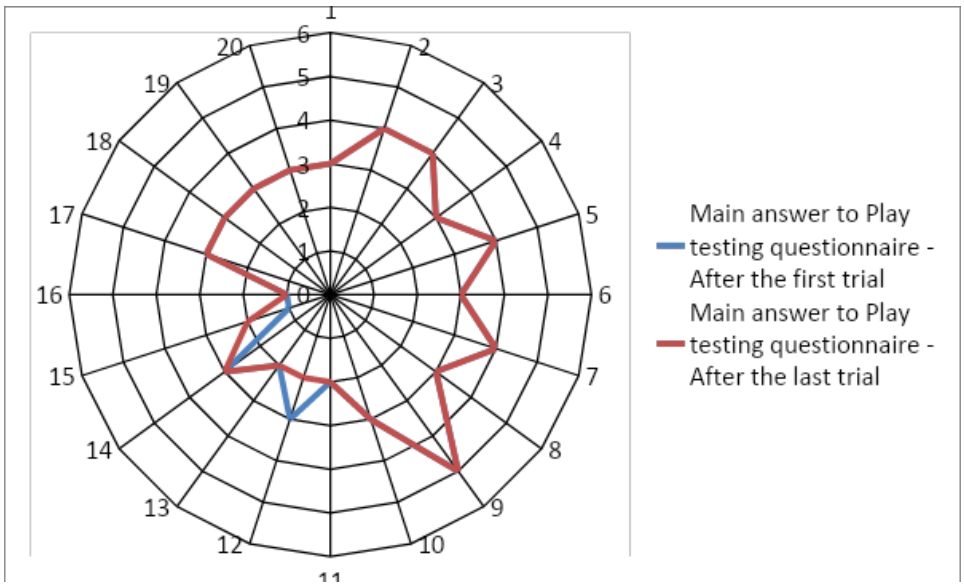
It is relevant to notice that for these users' items in **PO** are **relatively quite good** and also items in **NO** have **relatively low scores**. This can be explained saying that there could be correlation between MoCA test and how much usable they found the games: **the more cognitively capable they are, the easier the games become for them**.

Moreover, it is observed that **scores for NO items decrease** and the ones for **PO scores increase** in general (it is not true only for the seventh item from HTC Vive test).

In SUS for Oculus, using again a Wilcoxon test, it is seen that with a significance level **α = 0.1 items after first trial and after last trial are statistically different** for **items in PO and NO** respectively. On the other hand in SUS for HTC Vive, if a significance level **α = 0.1 is fixed**, it can be asserted that **items after first trial and after last trial are statistically different** for **items in NO**, but in this case it cannot be asserted that items after first trial and after last trial are statistically different for items in NO.

5.2.6. PTQ-ANA





Tables 27 & 28- PTQ results

	Item	First trial	Last trial
1	Brightness	just right	just right
2	3D effect	ok	ok
3	Models	nice	nice

4	Lighting	ok	ok
5	Game graphics overall opinion	nice	nice
6	Sound volume	well balanced	well balanced
7	Sound effects	nice	nice
8	Steering	just right	just right
9	Game targets	no problem	no problem
10	Game speed	just right	just right
11	Frying the egg	easy	easy
12	Introducing forms	just right	easy
13	Taking the money	easy	easy
14	Grabbing objects	just right	just right
15	Best device	Oculus	HTC Vive
16	Best Game	Forms	Forms
17	Forms	Pleasant	Pleasant
18	Bank	Pleasant	Pleasant

19	Kitchen HTC Vive	Pleasant	Pleasant
20	Kitchen Oculus	Pleasant	Pleasant

Table 29 - PTQ - ANA interpreted mean results

	HTC Vive			Oculus
	Forms	Kitchen	Bank	Kitchen
Attempts	89	88	86	94
Total time	3:41:39	1:15:48	1:51:33	2:11:46
Mean time	0:02:29	0:00:52	0:01:18	0:01:24

Table 30 - Attempts and timing

The first graphic shows the mean answer for each item after first trial (blue) and after last trial (red). The second one shows the main (most repeated) answer for each item after first trial (blue) and after last trial (red). It is clear that **both** graphics are **almost the same**; but taking the second one into account, it is seen that **users changed their preferences from Oculus to HTC Vive** (we can see it more clearly in the table). The rest is almost the same and it is interpreted in the table.

In general, all **items** are **well scored** and **games** are considered as **easy and pleasant**.

The **most played** game was **Forms**, which is the same that the best game voted in PTQANA.

5.2.7. Gender comparison

	n1	n2	U	U(n1,n2,0.05)
Interaction with new technologies	7	21	73,5	41
Usage of technologies	7	23	78	68
MoCA	8	25	62,8	60
EQ-5D-5L	7	25	70	50
GDS	8	25	86	60

Table 31 - U tests for gender comparison in Romania.

It can be considered if there is any difference between results depending on the gender. Four U test are done in order to know this. As it is shown, **it can not be concluded statistical differences to exist** at any relevant significance level for **any item.**

5.2.8. Correlation between MoCA and usability of games

	n1 (MoCA under 25)	n2 (MoCA equal or over 25)	U	U(n1,n2,0.05)
SUS Oculus	6	23	13	37

SUS HTC Vive	6	23	19,5	37
PTQ-ANA	7	23	36.5	46

Table 33 - U tests for tests comparison between patients under 25 in MoCA test and patients equal or over 25 coefficient for MoCA and these tests correlation.

	Entire group	MoCA score under 25
PTQ-ANA	-0.1093334	0.01961161
SUS Oculus	0.004641567	0.3134678
SUS HTC Vive	-0.2053562	0.3530939

Table 34 - Spearman's ρ

Again, first column represents Spearman's ρ coefficient for the entire group and the second one represents Spearman's ρ coefficient for patients whose MoCA score is under 25 points.

There are **statistical differences** between scores which belong to **patients who are under 25 in MoCA test and those which belong to the rest of them**. This happens for **every test**: SUS Oculus, SUS HTC Vive and PTQ-ANA.

Moreover, there are no significant correlation between variables.

5.3. Relationship between Denmark and Romania

The possible relationship between both groups can be studied. Attending to the health, the usage of technologies and MoCA results in both groups it is seen that they are quite different. As it is said before, thanks to the Mann-Whitney U-test it can be claimed that, with a significant level **α = 0.01**, both set of data are statistically different. This means that **health**, **technology usage** and **MoCA results** all **in Romania group are greater than in Denmark one.**

	n1 (men)	n2 (women)	U	U(n1,n2,0.05)
Interaction with new technologies	12	27	155,5	107
Usage of technologies	14	31	191,5	149
MoCA	16	36	212,5	204
EQ-5D-5L	11	30	151,5	108

Table 34 - U tests for gender comparison between men and women in Romania and Denmark

Finally, it can be considered if there is any difference between results depending on the gender.
Four U test are done in order to know this. As it is shown, **it can not be concluded statistical differences to exist** at any relevant significance level for **any item**.

Correlation between MoCA and usability of games

In this case results from SUS and MoCA tests have been considered together. It is shown that **ρ = 0.4423507**, if the entire group is considered. Even more, **ρ = 0.2823673** arises if patients under 25 in MoCA tests are considered. In conclusion, **there are no correlation between them in any cases**. (Spearman's ρ coefficient)

This time, as patients under 25 are 26 and the rest are 52, U test can be approximated by a normal distribution where:

$$\square = (\square - \square)/\square$$

$$\square = \square \cdot \square/2$$

$$= \sqrt{\frac{\square \cdot \square(\square + \square + 1)}{12}}$$

For this two groups, z = - 0,19079 which probability is 0,4247. This means that **it can not be concluded statistical differences to exist**.

6. Conclusions

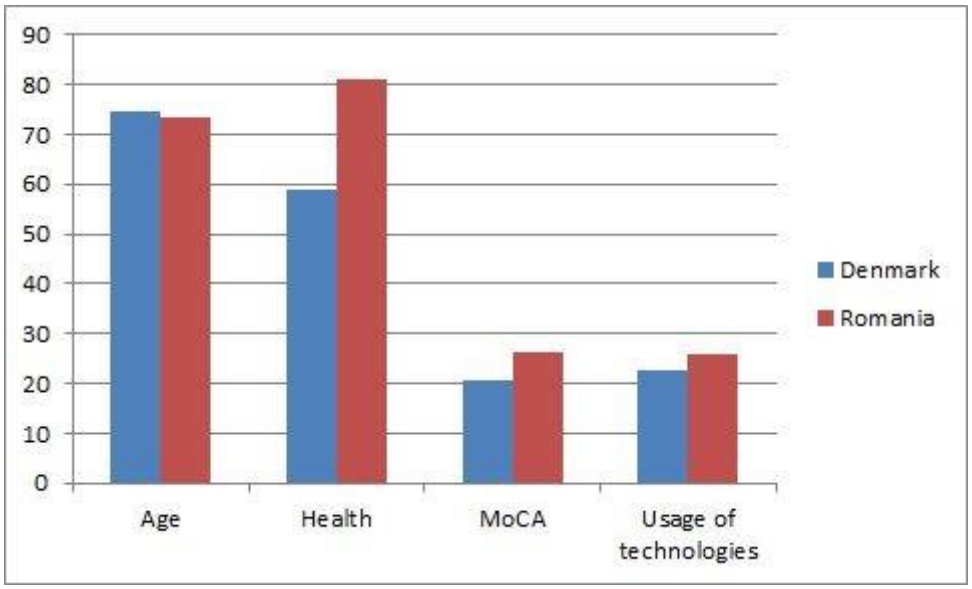
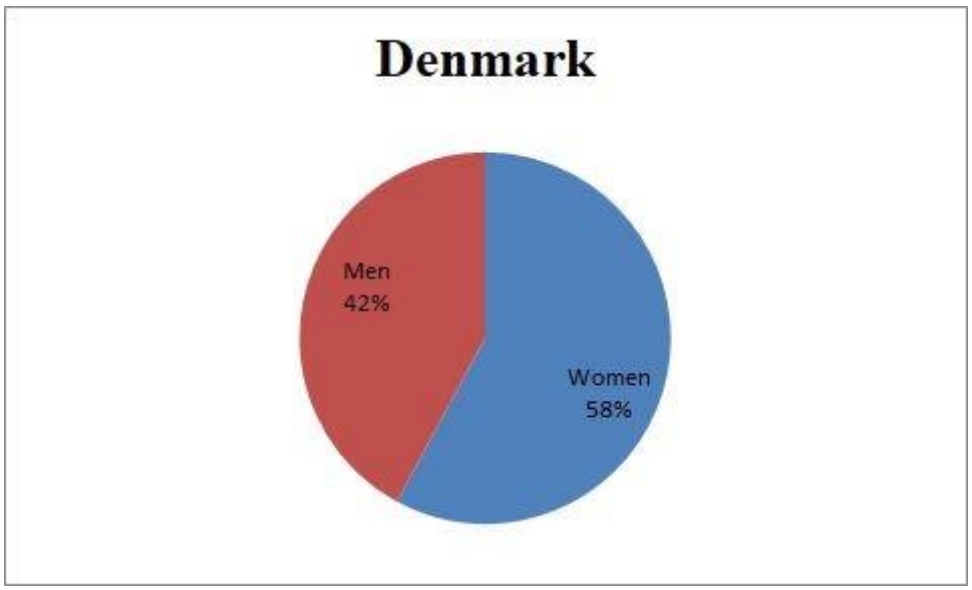
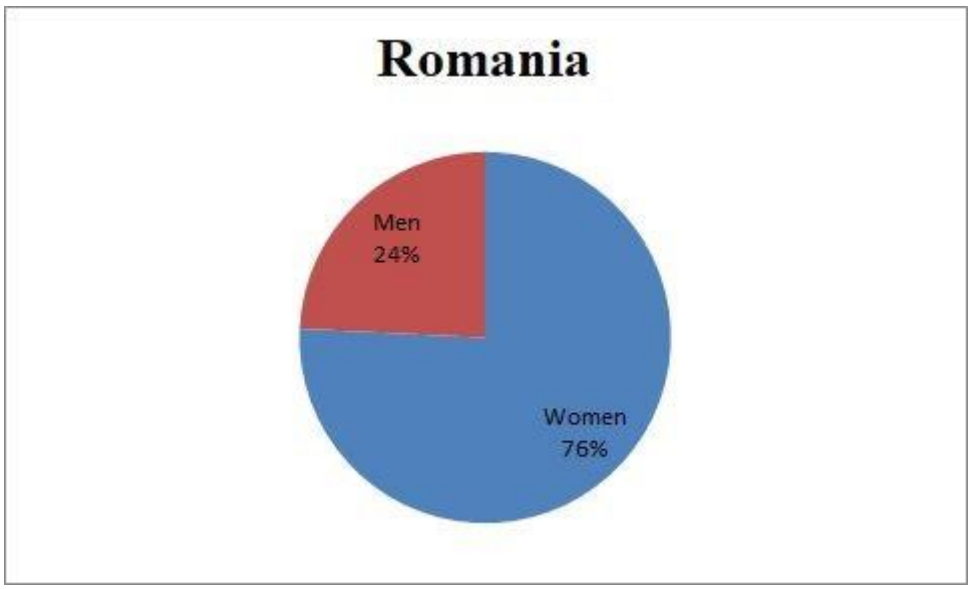


Table 29 - Denmark / Romania comparison





Tables 36 & 37 - Gender distribution in Denmark and Romania

Most important facts are:

Denmark:

- The ratio women/men is **57, 895%.**
- **Four educational** levels appeared the **same number of times**: 'high school',

  'professional school', 'superior studies' and 'other'. These others are: 'tailor (4 years education)', 'similar to nowadays 7th grade but in the old school system in Denmark' and 'have worked on a farm/housework'. - The most common living status is **living alone**.
- They got **22,5/55** points in **usage** of new technologies and it is **slightly-medium correlated** with their **interaction with new technologies**.
- They got **58, 75%** of mean health in **EQ-5L-5D**.
- **MoCA** test situated them **between 16 and 25**. - The most played game was **Kitchen**.

Romania:

- The ratio women/men is **75, 75%**.
- The most common educational level is **superior studies**.
- The most common living status is **living with spouse**.
- They got **26,5/55** points in **usage** of new technologies, but it is **not correlated** with their interaction with new technologies.
- They got **80, 9%** of mean health in **EQ-5L-5D**.

- **MoCA** test situated them **between 24 and 28**.
- **Nine people** are possible to have **depression**.
- **Games** are considered **easy** and **pleasant**.

- The most played game was **Forms**.

Summing up, there are **two very different groups** and that Romanian one had better health condition, so this could make them to consider the games to be easier than in Danish group whom integrant had more issues. Those issues made them need more assistance and then they could consider the games to be more difficult or just less usable.

Although the data related to the cognitive level does not show a great improvement, and maybe we lack more sessions, a control group, etc., the obtained results also indicate that the cognitive functions were maintained or even slightly improved during the project.

Results from qualitative analysis of the data:

Generally, equipment is not a problem for users but some of them had glasses and they were not able to use them, so they did not see as well as it was supposed to be.

There were also users who had to use a wheelchair and this made harder playing to Shapes game to them. That's why a wheelchair mode was created. Moreover, they have to take care of the wire of the equipment.

Some of them complained about the weight of HTC Vive headset.

In Kitchen, many users said some problems. First of all, the pan tends to fly around and it is hard to grab it. The second thing is that they can not turn it on with the left hand, and this is a problem for people who have an issue in their right arm. One more issue is that, when they fried eggs, many times eggs put one on top of each other and this is not realistic. The size of the written text supposed a problem too because they weren't able to see it. Moreover, both in Denmark and Romania complained about the language, because several things are only in English and in Denmark, there were translated things but there were some mistakes.

In Bank, some of them got angry because the ATM's keyboard was too high for them and also the money disappears very quickly so they were not able to catch it. Anyway, a control for adjusting height was included.

In Forms, sometimes the shapes flew away from the user and they were hard to grab too.

In general, users and therapists said that games were too easy and they wanted the games to let them increase the difficulty.

Almost every user did not feel any dizziness neither during the games neither after them.