User Satisfaction in Augmented Reality-based Training using Microsoft HoloLens

Hui Xue [1,*] **, Puneet Sharma** [1] **and Fridolin Wild** [2]

[1]   UiT Arctic University of Norway; hui.xue@uit.no; puneet.sharma@uit.no
[2]   Oxford Brookes University, United Kingdom; wild@brookes.ac.uk
*   Correspondence: UiT Arctic University of Norway; Tel.: +47-776-46-625

Abstract: With the recent developments in augmented reality (AR) technologies comes an increased interest in the use of smart glasses for hands-on training. Whether this interest is turned into market success or not depends at the least on whether the interaction with smart AR glasses satisfies users, an aspect of AR use that so far has received little attention. With this contribution, we seek to change this. The objective of the article, therefore, is to investigate user satisfaction in AR applied to three cases of practical use. User satisfaction of AR can be broken down into satisfaction with the interaction and satisfaction with the delivery device. A total of 142 participants from three different industrial sectors contributed to this study, namely, aeronautics, medicine, and astronautics. In our analysis, we investigated the influence of different factors, such as age, gender, level of education, level of Internet knowledge, and the roles of the participants in the different sectors. Even though users were not familiar with the smart glasses, results show that general computer knowledge has a positive effect on user satisfaction. Further analysis using two-factor interactions shows that there is no significant interaction between the different factors and user satisfaction. The results of the study affirm that the questionnaires developed for user satisfaction of smart glasses and the AR application performed well, but leave room for improvement.

Keywords: augmented reality; Microsoft HoloLens; AR application; user experience; user satisfaction

1. Introduction

Augmented Reality (AR) means enhancing the user's perception "with additional, artificially generated sensory input to create a new experience including, but not restricted to, enhancing human vision by combining natural with digital offers" (Wild et al., 2018). Augmented Reality typically has three characteristics [1]: first, AR combines the virtual with the real world; second, objects are registered from both the real and virtual world in one coordinate system; third, the interaction between the objects of both worlds is possible in real time.

Hands-on training is important for many disciplines and professions, such as medical workers, mechanics, technicians, electricians, engineers, sailors, pilots, and firefighters. In the past decade, AR has been increasingly employed for a number of training applications, such as medical education [2], rehabilitation engineering [3], automotive safety [4], task assistance [5], and manufacturing [6].

For the successful adoption of AR-based training across different domains, one of the key factors is user satisfaction. User satisfaction is defined as a combination of different factors associated with the usage of the AR application and the associated delivery device[7]. These factors include: a feeling of powerfulness and achievement; an efficient use of time, effort, and other resources; meaningful content; a better insight to the training environment; a natural interaction; a feeling of amazement; performance that exceeds expectations; playfulness; the invoking of positive feelings and pleasing memories; immersion and engagement; a transparent interaction; the feeling of participation in a community; a sense of privacy of the user's content; inspiration, encouragement, and motivation; and, finally, artistic creativity [7].

The rest of this paper is organized as follows. First, we turn to the state of the art, summarizing what the research has found so far with respect to AR user interaction, AR user satisfaction, and questionnaires used for evaluating user satisfaction. Next, the AR app used in the trials is described. Subsequently presented are the research methodology and a summary of the information of the participants, devices, design of trial tasks, and evaluation methods. Finally, findings and results are illustrated, and the discussion and conclusion are given at the end.

The main objective of this study is to test and observe user satisfaction in using AR applications and using AR glasses. The method for evaluating includes questionnaires and interviews. The AR app used in this evaluation, therefore, has two parts: one is the expert recording the experience in the workplace, and the other part is the novices training on work-related procedures using said recordings. In this study, we evaluated the following research hypotheses: to find if experts and students are satisfied with the prototype application, to see if the application can increase interest in learning new skills, and to evaluate if the users find the application easy to use.

2. State of the art

2.1. AR user interaction

AR technologies provide a different user experience than that of, for example, mobile phone apps. The user interacts with the surrounding real world, combining inputs from the environment with digital augmentations. Popular examples include PokemonGO and SnapChat. These type of apps certainly brought the term "augmented reality" into the spotlight [8]. With the advent of consumer-grade AR glasses, different types of AR user interactions are becoming necessary. For example, a user who is wearing Microsoft's HoloLens can communicate diagrams and other types of graphics directly embedded into the environment to a different, remote user (see Figure 1).
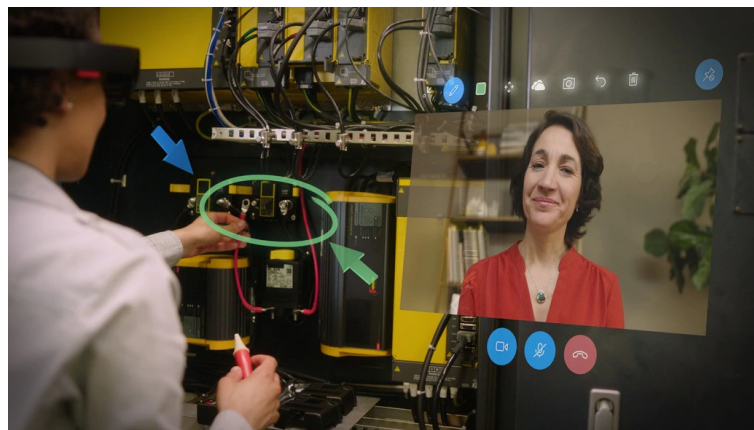
Figure 1. With Microsoft HoloLens, a user connects the wires with remote assist [9].

2.2. AR user satisfaction and questionnaires for evaluating user satisfaction

AR Technology has evolved from offline to online, from static devices to mobile devices, and from desktop and mobile to wearable devices [10]. Consequently, with AR development over the past decade or so, special attention has been drawn to the maximization of AR user satisfaction. AR user satisfaction is dependent on both the design of the user interface (UI) and the choice of the AR hardware. Personalization of AR glasses can lead to greater AR user satisfaction [11]. AR apps designed for a good user experience result in a more overall satisfied AR user. This applies to AR navigation apps, AR health apps, AR education apps, and certain AR smart glasses games [12].

There are several concepts and subjective measures for evaluating the user experience of AR services. With regards to the user, satisfaction questionnaires are common tools used to evaluate a user's experience. One such tool—the Questionnaire for User Interaction Satisfaction (QUIS)—is

Figure 2. User Interface of the recording mode. Image from the WEKIT consortium in 2017.

(a)   (b)
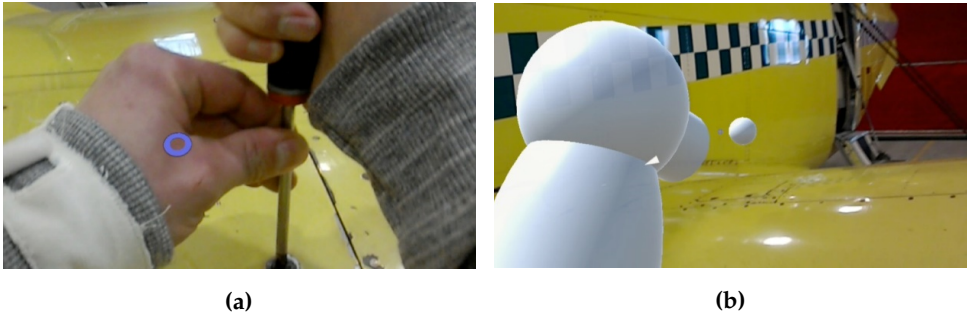
Figure 3. A ghost track in WEKIT Recorder mode: (a) recording a ghost track; (b) ghost track replay. Image from [17].

designed to assess users' subjective satisfaction with specific aspects of the human–computer interface [13]. The results of QUIS facilitate new developments by addressing reliability and validity problems found using its satisfaction measurements. Therefore, the measure is highly reliable across many types of interfaces.

QUIS consists of a demographic questionnaire, a six-scale measure of overall system satisfaction, and hierarchically organized measures. The measures include the following specific interface factors [13]: screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, online tutorials, multimedia, teleconferencing, and software installation. Each area is measured by a 7-point scale according to the user's overall satisfaction with the interface and the above factors [13].

3. The AR application

The trials of the project investigated how satisfied users are with the novel method and the AR glasses. The app is designed for HoloLens with two modes. One is called the WEKIT (Wearable Experience for Knowledge Intensive Training) recorder and another one is called the WEKIT player. The recorder mode tracks and records the performance of the experts. To create the required instruction for a procedure, experts can create annotations for each action step of the procedure. These annotations can then be played back to the trainees. There are several types of annotations that can be added to this app. Figure 2 is the user interface (UI) of the recording application. Each icon represents a different type of annotation. Figure 3 shows a ghost track recording and replay.

The WEKIT player is the mode designed for trainees to learn the operation that was recorded before. Each scenario has a different recording and tasks. Therefore, the app starts by recognizing an Augmented Reality marker, and then it decides which scenario is going to be used in the subsequent tasks. Markers are always detected by the front camera on HoloLens. The WEKIT player starting screen is shown in Figure 4. Once the task starts, the annotation is shown in HoloLens. From the perspective of the users, the annotations overlap with the facilities. They guide the user to do the task step by step. Gesture command, voice command, and the Physical HoloLens click button are all available when using the app. Figure 5 shows an example of using the WEKIT player to do the task.
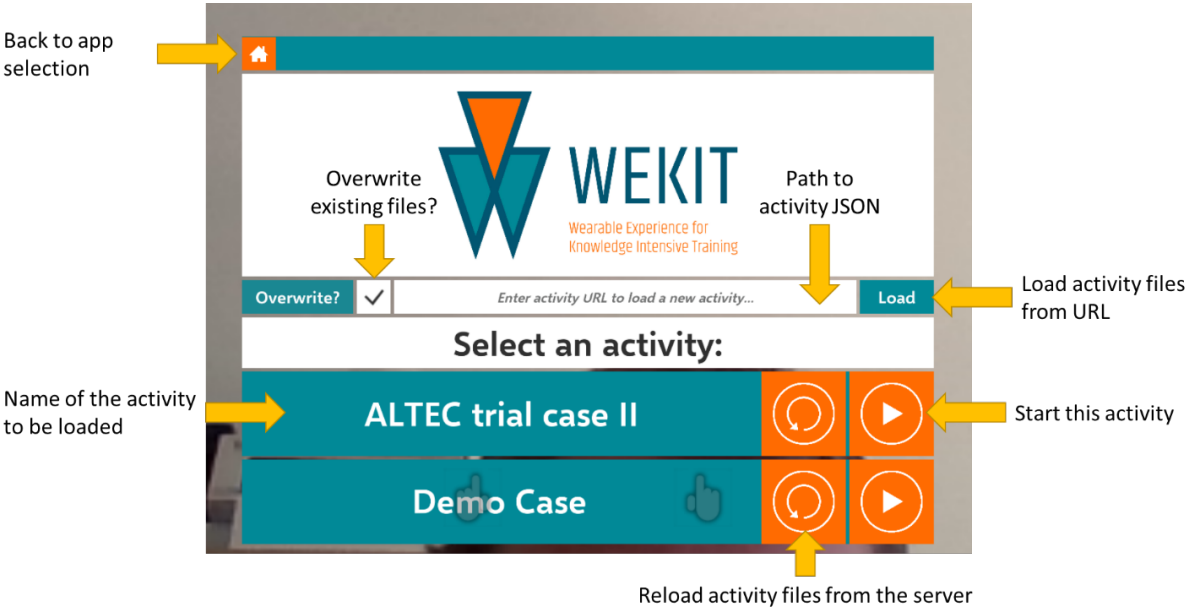
Figure 4. Starting screen in WEKIT Player mode. Image from [15].

Figure 5. Example of user interface of WEKIT Player mode. Image from [15].

4. Research Design/Experiment Methodology

4.1. Participants

In order to evaluate the satisfaction of the user's interaction and the smart glasses user experience, the WEKIT application was designed for three different use cases: aviation, medical imaging, and space. In our trial experiments, the test population was divided into two main groups: experts and students. A total of 47 experts (8 females; 39 males) with a high level of technological competency in

their respective fields were recruited. A total of 95 learners (23 females; 72 males) from different fields, including medicine, engineering, and aerospace, voluntarily participated in the trials. A majority of the participants (68) were in the 18–24 age group, followed by 48 of the participants in the range between 25 and 34. Most of the participants had moderate or better computer knowledge and internet knowledge. Here, we defined computer knowledge and internet knowledge as very poor, poor, moderate, good, and very good for five different levels. All participants gave written consent for the trials.

4.2. Material and Apparatus

The trial used Microsoft HoloLens as wearable AR glasses for assessing the user's satisfaction with AR training. There are two parts in the WEKIT technology platform [16] deployed on HoloLens. One is a recorder for capturing expert experience and the other one is a player for presenting the expert's experience to the trainees. During the trial, all interactions and manipulations were done by using gesture and voice command only.

4.3. Trial design/task

The trial tasks were separated into three different areas, as mentioned in section 4.1. Tasks in the Aeronautics use case were performed at Lufttransport, Norway. A pre-flight inspection task was performed with the air ambulance plane Beechcraft B200 [17]. The experts comprised maintenance apprentices, skilled workers (mechanics), and technicians working on base maintenance at Lufttransport. As an example, an expert wearing HoloLens and using the application to record their voice, image, and movement is shown in Figure 6. The novice group comprised student volunteers from UiT The Arctic University of Norway [17]. They followed the instructions from the application on HoloLens for completing the task.
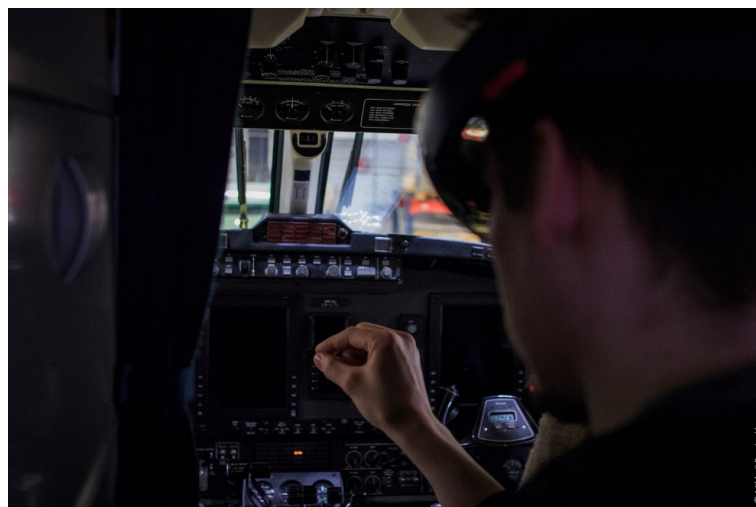
Figure 6. Maintenance Engineer in the cockpit of a Beechcraft B200 King Air model. Image photographed by Mikhail Fominykh in 2017.

The medical task involved imaging and diagnostic workers and was conducted at EBIT (Esaote's Healthcare IT Company) in Genoa, Italy [18]. This task was for training medical students and radiologist apprentices on using MyLab8, an ultrasound machine produced by ESAOTE [19]. Similar to the trial at Lufttransport, the experts added audio recordings, pictures, annotations, and 3D models by using the recorder application. The novices performed the task based on the recorded expert's experience. In Figure 7, we can see a novice performing a task by positioning the probe in the target direction and taking measurements using the player application.

Figure 7. A radiologist conducting ultrasound training. Image from the WEKIT consortium in 2017.

The space task that was conducted at the ALTEC facility in Turin involved training astronauts to install the Temporary Stowage Rack (TSR). TSR installation is a procedure that the astronauts have to perform on the International Space Station (ISS) [20]. Similar to the trials at the other two organizations, the experts designed the training scenario and added annotations by using the recorder application. The novices performed the task based on the recorded content. In Figure 8, we can see an astronaut trainer performing a task in a replica-training module of the international space station.
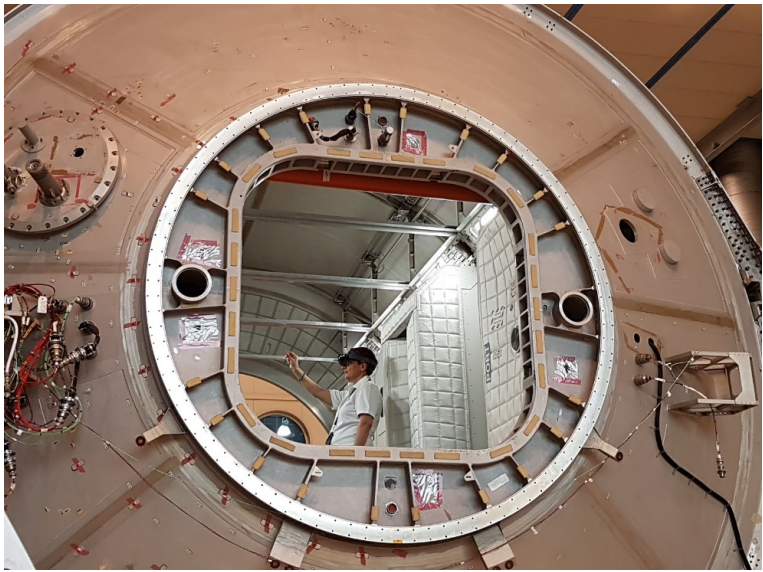
Figure 8. Astronaut trainer in a replica-training module of the international space station. Image from the WEKIT consortium in 2017.

4.4. Smart Glasses User Satisfaction (SGUS)

The Smart Glasses User Satisfaction (SGUS) questionnaire was created for the WEKIT trials. It is a tool designed to assess users' subjective satisfaction with smart glasses. SGUS is a method and measure to scrutinize aspects, such as an enhanced perception of the environment, interaction with the augmented environment, implications of location and object awareness, the user-created AR content, and the new AR features that users typically use [7]. The general objective of the questionnaire is to understand the potential end users' central expectations of AR services with smart glasses, especially from an experiential point of view [7]. In this study, the smart glasses used for the different use cases were Microsoft HoloLens. SGUS measures subjective satisfaction on the basis of different features associated with user satisfaction, such as the content and interaction with the content. SGUS is based

147 on evaluation criteria for web-based learning [14] and statements evaluating the user experience of
148 mobile augmented reality services [7]. SGUS consists of 11 items (statements) on a 7-point Likert scale
149 (1–7) [17]. The 11 statements include three categories of evaluation criteria, which are general interface
150 usability criteria, AR interaction-specific criteria for an educational AR app, and learner-centered
151 effective learning [14].

152 *4.5. Questionnaire for User Interface Satisfaction (QUIS)*

153   The Questionnaire for User Interaction Satisfaction (QUIS) measures subjective satisfaction with
154 specific aspects of the interface and interaction between the user and the AR application [21]. In this
155 study, QUIS was modified for AR glasses, i.e., HoloLens. Hence, a questionnaire with 15 items was
156 used. In order to maintain consistency with the survey in other sections, each item was mapped to a
157 numeric value of 1–7 instead of the 9-point scale [21].

158 *4.6. Procedure*

159   As most participants had no experience with AR glasses, at the beginning of the trial, they were
160 asked to familiarize themselves with the AR glasses, i.e., HoloLens. In order to do this, gesture training
161 with HoloLens was done before they started using the application. The application comprised a
162 scenario that the participants had to complete in a particular use case setting. The content of the
163 application was generated by experts in that specific use. After the participants completed all the tasks,
164 they were provided with the QUIS and SGUS questionnaires to complete.

5. Results/Findings

166 *5.1. Descriptive statistics*

167   In this section, descriptive statistics for SGUS and QUIS are described.

168 5.1.1. SGUS

169   As mentioned before, SGUS has 11 items. The summation of the score for the 11 items is the
170 SGUS score. As shown in Table 1, we provide data such as: n (number of participants), mean, standard
171 deviation, minimum value, Q1 (the first quartile: "middle" value in the first half of the rank-ordered
172 data set), median, Q3 (the third quartile: "middle" value in the second half of the rank-ordered data
173 set), and maximum value for the following variables: gender, education level, roles, and organizations.
174 Based on these results, it is clear that the mean scores are similar across the different levels associated
175 with the variables.

Table 1. Descriptive statistics of the Questionnaire for Smart Glasses User Satisfaction (SGUS).

Variable	Level	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
Gender	Female	31	58.74	7.96	43	54.5	58	64.5	72
	Male	111	58.49	8.45	20	54	60	64.5	72
Role	Experts	47	56.98	8.83	33	49.5	58	64	72
	Students	95	59.32	7.99	20	55	60	65	74
Education level	Upper secondary school or lower	45	57.98	7.92	33	54	57	64	70
	Bachelor's or higher	97	58.8	8.52	20	55	60	65	74
Organization	Space(1)	39	59.54	9.46	20	57	61	65.5	71
	Medicine(2)	48	58.69	7.43	38	54.75	59	64	72
	Engineering(3)	55	57.71	8.26	33	52	57	64	74

176 5.1.2. QUIS

177 Similarly, the overall Questionnaire for User Interface Satisfaction (QUIS) score was calculated by
178 summation of the score for the 15 QUIS items. Summary data for all questions in QUIS are presented
179 in Table 2. The 15 items were designed independently from each other. These items aim to investigate
180 the satisfaction of users with different aspects of the interface, including usability and user experience
181 in using AR applications.

Table 2. Descriptive statistic of the Questionnaire for Smart Glasses User Satisfaction (QUIS).

Variable	Level	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
Gender	Female	30	75.94	11.44	55	69	76.5	82.75	98
	Male	103	76.99	13.23	18	71	78	86	103
Role	Experts	43	76.28	12.06	49	69	77	85.5	97
	Students	90	77.01	13.21	18	72	78	85	103
Education level	Upper secondary school or lower	43	75.14	13.82	18	69.5	75	85	95
	Bachelor's or higher	90	77.56	12.30	33	71	78	85.75	103
Organization	Space(1)	39	76.67	12.44	33	72	77	86	96
	Medicine(2)	42	80.50	9.71	55	75	80	85.75	97
	Engineering(3)	52	73.85	14.61	18	66	74.5	83.5	103

182 *5.2. Correlation*

183 In this section, we discuss correlation for SGUS and correlation for QUIS.

184 5.2.1. Correlation of SGUS

185 Spearman's correlation coefficient, ρ, measures the strength and direction of association between
186 two ranked variables in the range [-1, 1]. Based on the 11 items, the results of Spearman's rank
187 correlation are shown in Table 3: the first value of each row represents Spearman's correlation
188 coefficient, and the second value of each row represents the p value. It can be seen that almost
189 all items are statistically significant ($p < 0.05$) and have a low positive correlation. This implies that all
190 the items are independent.

191 In the study of SGUS, each of the items investigates a different aspect of the user experience. For
192 the analysis, the overall averages for all items were calculated. Figure 9 shows the plot of the average
193 score from individual items. The box in the plot depicts the answer of 50% of the participants, with the
194 line in the middle indicating the median. The dotted lines span the 95% confidence interval. Outliers
195 are depicted with black dots. The connected red dots indicate the medians. The results imply that
196 most of the participants had a good conception of what is real and what is augmented when using
197 AR-glasses (GL5). The participants indicated that the system and content helped them to accomplish
198 the task quite well (GL7) and their attention was captivated in a positive way (GL6). The provided
199 content was also seen as contextually meaningful (GL2). However, performing the task with AR
200 glasses was experienced as less natural (GL9, GL4), and following and understanding the task phases
201 (GL8, GL10–11) was not very easy. The results were very much in line across the three.

Table 3. Spearman's rank coefficient of correlation for SGUS: the first value of each row represents Spearman's correlation coefficient, and the second value of each row represents the p value.

	GL1	GL2	GL3	GL4	GL5	GL6	GL7	GL8	GL9	GL10	GL11
GL1	1	0.316 0.000**	0.209 0.013	0.269 0.001	0.164 0.053	0.301 0.000**	0.270 0.001	0.323 0.000**	0.285 0.001	0.376 0.000**	0.336 0.000**
GL2		1	0.335 0.000**	0.371 0.000**	0.239 0.005	0.308 0.000	0.345 0.000**	0.227 0.007	0.287 0.001	0.354 0.000**	0.398 0.000**
GL3			1	0.487 0.000**	0.172 0.041	0.270 0.001	0.444 0.000**	0.312 0.000**	0.320 0.000**	0.265 0.002	0.289 0.001
GL4				1	0.121 0.154	0.293 0.000**	0.376 0.000**	0.337 0.000**	0.492 0.000**	0.226 0.008	0.243 0.004
GL5					1	0.260 0.002	0.178 0.036	0.170 0.046	0.026 0.763	0.062 0.468	0.166 0.052
GL6						1	0.416 0.000**	0.387 0.000**	0.453 0.000**	0.317 0.000	0.403 0.000**
GL7							1	0.490 0.000**	0.500 0.000**	0.453 0.000**	0.435 0.000**
GL8								1	0.563 0.000**	0.442 0.000**	0.390 0.000**
GL9									1	0.455 0.000**	0.364 0.000**
GL10										1	0.558 0.000**
GL11											1

Signif. codes: *p < 0.05 ; **p < 0.01; ***p < 0.001

Smart Glasses User Satisfaction (All)
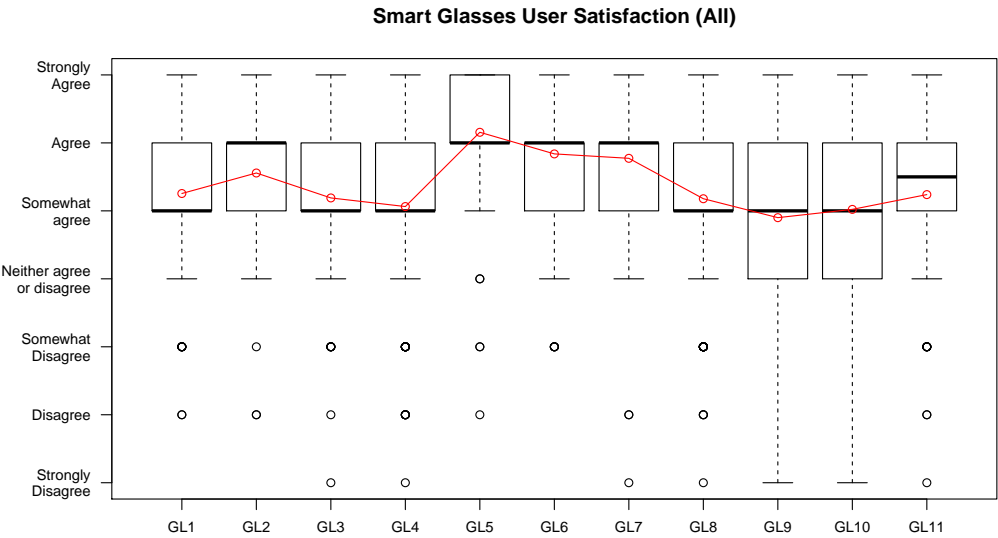
Figure 9. Plot of SGUS score for each item.

5.2.2. Correlation of QUIS

The correlation for QUIS is based on 15 items. The results of Spearman's rank correlation are shown in Table A1 (see Appendix). The values in the table have the same meaning as in Table 3. The results are similar to those of SGUS; most of the items are statistically significant (p < 0.05) and have a low positive correlation. This implies that most of the items are independent.

In the study of QUIS, each of the items investigated different aspects of the user experience. For the analysis, the overall average from all items was calculated. Figure 10 shows the plot of the average score from individual items, and the description of the plot is the same as that of the SGUS plot. The results imply that most of the participants agree that learning to operate the AR glasses (QS13) seemed

to be rather easy, and the overall enthusiasm towards the system seemed (QS1, QS5) to be very positive. The characters on the screen were relatively easy to read (QS9). The means of QS3, 4, 6, 7, and 8 indicate that the system was experienced as rigid, unreliable, and slow, which may cause frustration [17].
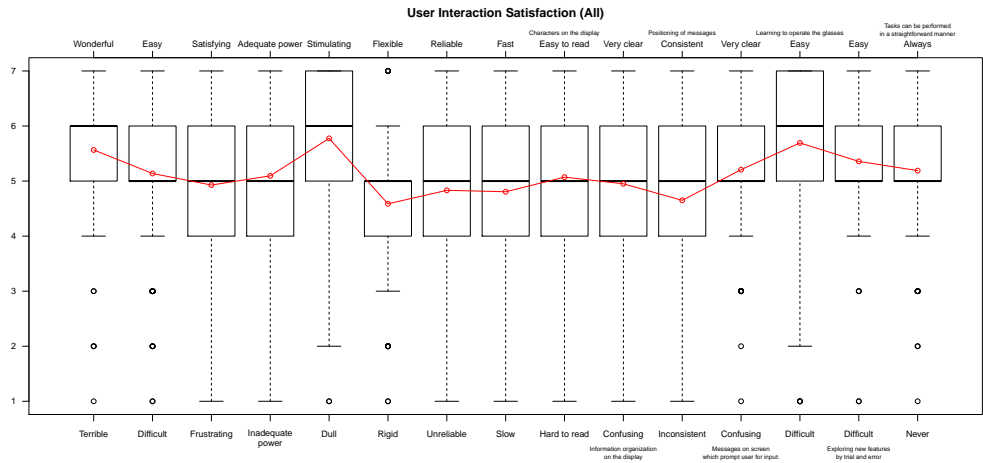
Figure 10. Plot of QUIS score for each item.

5.3. Analysis of variance and Interaction plots

The participants are described by seven factors: gender, age, role, education skill level, computer knowledge level, internet knowledge level, and organization. Each factor is divided by two levels, except for organizations, which are in three levels. Please note that none of the participants claimed that they have a poor or very poor internet knowledge level. The following section discusses the analysis of variance (ANOVA) of QUIS and of SGUS. In this ANOVA study, SGUS and QUIS scores were investigated for using the application on the AR glasses with six independent variables, i.e., the relationships between: age distribution, gender, roles, highest level of education, organization, and computer knowledge. Therefore, there are 6 main effects and 57 interactions. We are interested in whether there is a relationship between the satisfaction levels (measured by the questionnaire) and these factors.

5.3.1. ANOVA of SGUS

In this study, we investigated whether the age, gender, roles, computer knowledge level, or different organizations have an effect on the satisfaction of using AR glasses. To determine this, we needed to look at the simple main effects: the main effect of one independent variable (e.g., age) at each level of another independent variable (e.g., for students and for experts).

Figure 11 shows the main effects of the six factors. Participants with different computer knowledge levels have the greatest differences in the SGUS results. This means that the participants with good computer knowledge and poor computer knowledge gave different scores for user satisfaction. The results show that participants with good or very good computer knowledge were, in general, more satisfied with the smart glasses application, and there is a significant effect from computer knowledge levels (F value = 8.87, p = 0.003). The result implies that the SGUS score was affected by the effects of good computer knowledge.

Table 4 shows the summary results of the linear model of the independent variables. The estimate for the model intercept is 54.688 and the coefficient measuring the slope of the relationship with computer knowledge level is 4.324. There is strong evidence that the significance of the model coefficient is significantly different from zero: as the computer skill level increases, so does the satisfaction. The information about the standard errors of these estimates is also provided in the

Coefficients table. In the result of the multiple regression model, only 8.8% of the variance in the SGUS scores is explained by each of the factors (Multiple R-squared is 0.088). There is no statistically significant factor that explains the variation in the SGUS scores (overall p value is 0.08).
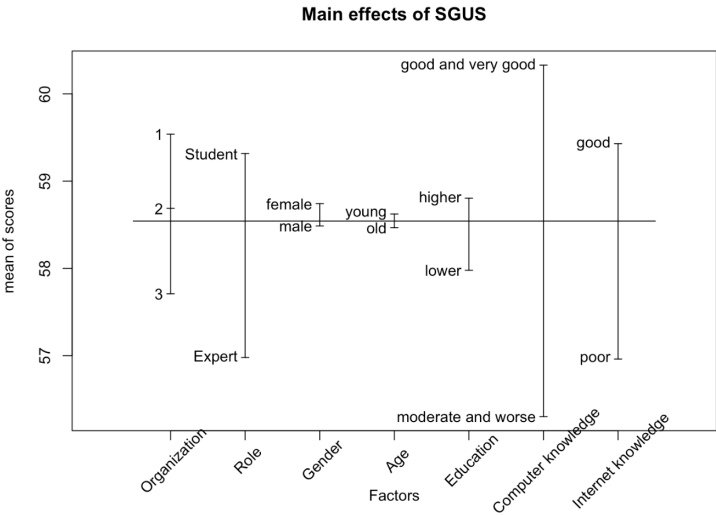
Figure 11. Main effects of SGUS.

Table 4. Results of the linear model of the independent variables.

Source of Variation	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	54.688	2.652	20.622	<2e-16***
Medicine	-1.604	1.835	-0.874	0.384
Engineering	-0.996	1.906	-0.523	0.602
Role	2.862	1.624	1.762	0.080
Gender	1.250	1.756	0.712	0.478
Age	0.563	1.634	0.344	0.731
Education level	-0.147	1.716	-0.086	0.932
Computer skill	4.324	1.452	2.978	0.003

Signif. codes: *p < 0.05 ; **p < 0.01; ***p < 0.001.

To investigate the interaction, it is interesting to find out whether the SGUS score depends on an interaction between good computer knowledge and the other factors. The two-factor interaction plot is shown in Figure 12. The following are the findings from the plot:

- Female participants with good computer knowledge have a higher SGUS score than males with good computer knowledge; both females and males with moderate and worse computer knowledge have nearly the same, lower SGUS score (Figure 12a).
- Participants from medicine with good computer knowledge tended toward a higher SGUS score than participants from engineering, and there is no significant difference between them and the participants with good computer knowledge from astronautics and medicine (Figure 12b).
- There is no significant interaction between participants with different computer knowledge levels from astronautics and engineering (Figure 12b).
- There is no significant interaction between students and experts with different computer knowledge levels (Figure 12c).
- Participants younger than 25 years old with good computer knowledge tended toward a higher SGUS than participants older than 25 years old; however, participants younger than 25 years old with a moderate and worse computer knowledge level tended toward a lower SGUS score (Figure 12d).

- Participants with secondary school or lower education level and good computer knowledge tended toward a higher SGUS score than participants with a bachelor's or higher education level and good computer knowledge level. However, participants with secondary school or lower education level and moderate and worse computer knowledge tended toward a lower SGUS score than participants with a bachelor's or higher education level and moderate and worse computer knowledge level (Figure 12e).

From the result of the ANOVA table (Table 5), there is insufficient evidence of statistical significance for two-factor interactions, since all p values are higher than 0.05.

Table 5. ANOVA results for SGUS with regard to organization, role, and computer knowledge level (reducing factors).

Source of Variation	Df	Sum Sq	Mean Sq	F value	Pr(>|F|)
Organization	2	77·9	38·95	0·576	0·563
Role	1	184·4	184·39	2·729	0·101
Gender	1	0·2	0·19	0·003	0·958
Age	1	4·2	4·16	0·062	0·805
Education level	1	0·0	0·02	0·000	0·988
Computer knowledge	1	589·3	589·31	8·723	0·004 ∗∗
Education level : Computer knowledge	1	65·0	64·98	0·962	0·329
Gender : Computer skill	1	121·5	121·49	1·798	0·182
Organization : Computer knowledge	2	28·9	14·47	0·214	0·807
Age : Computer knowledge	1	11·6	11·60	0·172	0·679
Roles : Computer knowledge	1	28·6	28·55	0·423	0·517
Residuals	128	8647·7	67·56		

Signif. codes: *p < 0.05 ; **p < 0.01; ***p < 0.001.

5.3.2. ANOVA of QUIS

In this section, the effect of the six independent variables (age, gender, roles, computer knowledge level, and different organizations) on user interaction satisfaction is reported. Satisfaction includes specific aspects of the interface, usability, and user experience of the AR application.

A total of 133 participants were used for this part of the study and completed the questionnaire. The simple main effects are shown in Figure 13. The results obtained by using the ANOVA in Table 6 indicate that the significance of the two-factor interaction of computer knowledge levels and organizations is not supported since all p values are more than 0.05. Table 6 also shows that the computer knowledge levels and different organizations have a significant effect on QUIS (p value is 0.008 for computer knowledge levels and 0.041 for different organizations).
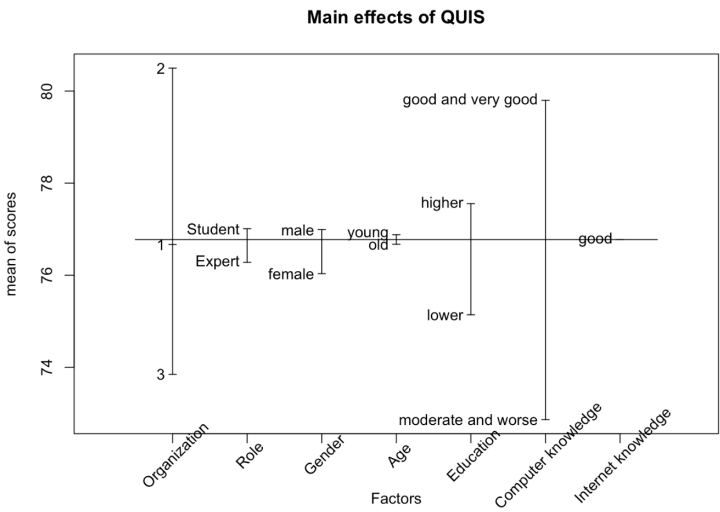
Figure 13. Main effects of QUIS.
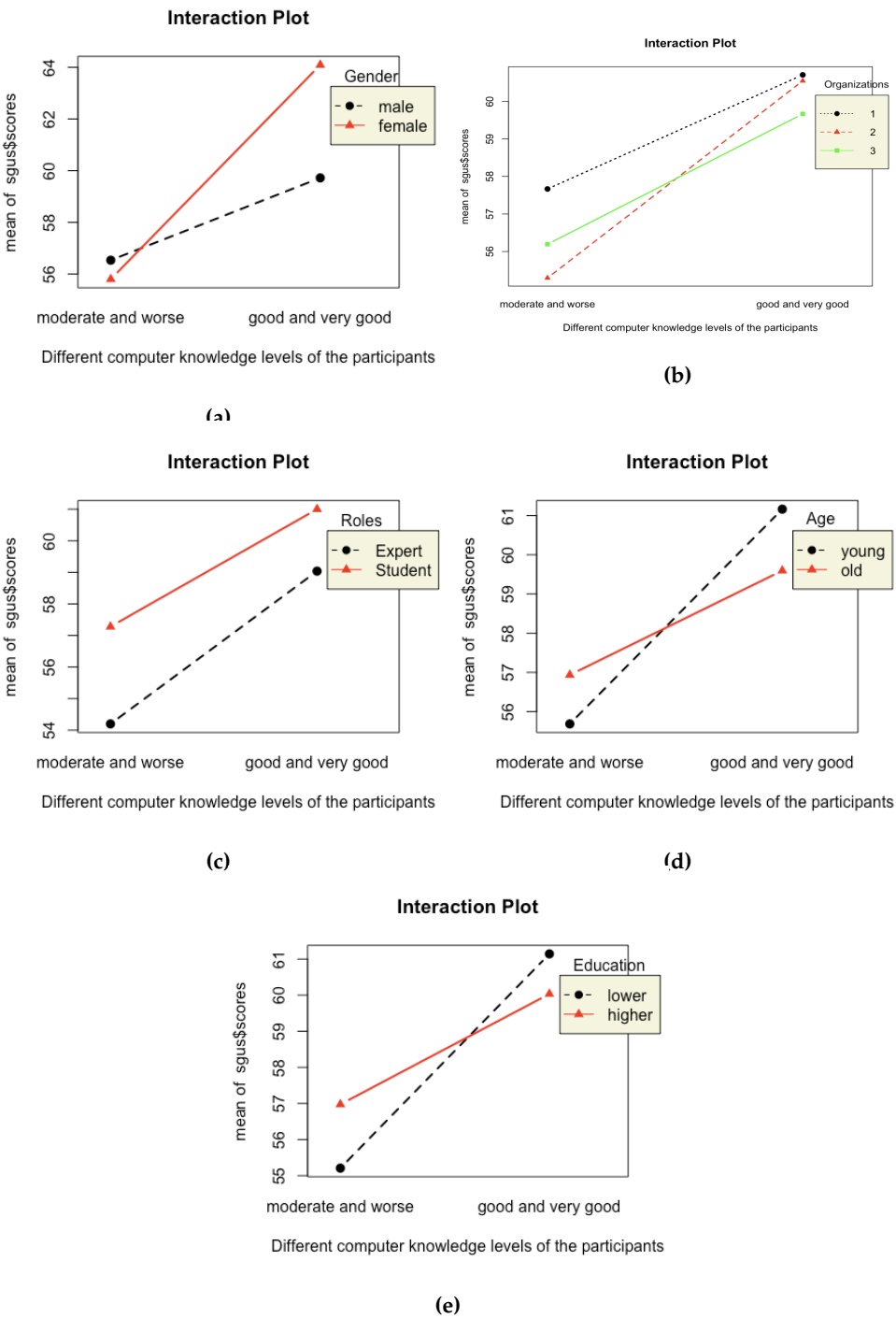
Figure 12. Interaction effects plots for SGUS: (a) Different computer knowledge levels with different genders of the participants; (b) Different computer knowledge levels with different organizations of the participants; (c) Different computer knowledge levels with different roles of the participants; (d) Different computer knowledge levels with different age groups of the participants; (e) Different computer knowledge levels with different education levels of the participants.

Table 6. ANOVA results for QUIS with regard to organization, role, and computer knowledge level (reducing factors).

Source of Variation	Df	Sum Sq	Mean Sq	F value	Pr(>|F|)
Organization	2	1029·3	514·65	3·279	0·041*
Role	1	10·4	10·37	0·066	0·798
Gender	1	90·3	90·31	0·575	0·450
Age	1	5·8	5·79	0·037	0·848
Education level	1	32·0	32·02	0·204	0·652
Computer knowledge	1	1138·1	1138·14	7·251	0·008**
Education level : Computer knowledge	1	165·5	165·55	1·055	0·307
Gender : Computer skill	1	449·7	449·74	2·865	0·093
Organization : Computer knowledge	2	0·9	0·46	0·003	0·997
Age : Computer knowledge	1	28·2	28·18	0·180	0·673
Roles : Computer knowledge	1	31·8	31·84	0·203	0·653
Residuals	119	18679·1	156·97		

Signif. codes: *p < 0.05 ; **p < 0.01; ***p < 0.001.
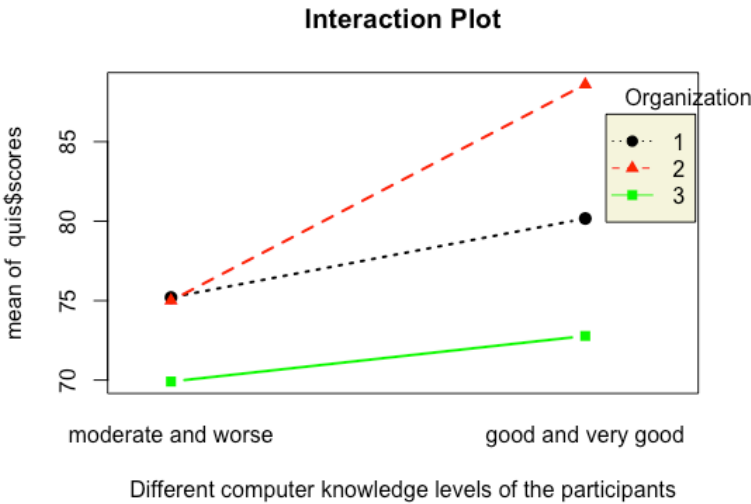
Figure 14. Interaction plot of different computer knowledge levels and the different organizations for QUIS

Figure 14 shows that in all three organizations, participants with moderate or worse computer levels were given lower scores than participants with good and very good computer levels. There are no significant interactions between them.

We selected the factors of organization and computer knowledge level to investigate the interaction between them, and the summary results of the linear model regression (see Table 7) shows that the estimate for the model intercept is 73.533, while there is no significant interaction between them. The information about the standard errors of these estimates is also provided in the coefficients table (Table 7). From the result of the multiple regression model, 10.6% of the variance in QUIS scores is explained by each of the factors (Multiple R-squared is 0.106). There is a statistically significant factor to explain the variation in the QUIS scores (overall p value is 0.0133).

Table 7. Summary results of the linear model of the independent variables for QUIS.

Source of Variation	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	73.533	3.188	23.063	<2e-16***
Medicine	2.533	4.509	0.562	0.575
Engineering	-2.748	3.951	-0.695	0.488
Computer knowledge	5.092	4.064	1.253	0.213
Medicine: Computer knowledge	1.805	5.686	0.317	0.751
Engineering: Computer knowledge	1.539	5.322	0.289	0.773

Signif. codes: *p < 0.05 ; **p < 0.01; ***p < 0.001.

6. Discussion

This study established a set of norms to be used for the evaluation of satisfaction of using AR glasses and AR applications. The relationship between each questionnaire item shows weak correlation, both in SGUS and in QUIS. Each questionnaire item is designed for evaluating a specific aspect of satisfaction of the smart glasses and AR applications. From the mean score of both questionnaires, we observe that most of the participants are satisfied with the AR glasses and the AR applications. It was found that the system and content helped the participants to accomplish the task quite well and their attention was captivated in a positive way. In other words, the result shows that the user interface is well designed. The user sees "useful information" displayed next to each part.

The main factors age, gender, education level, roles of the participants, and organizations do not have significant effects on the satisfaction of using smart glasses and AR applications. However, computer/internet knowledge level does influence user satisfaction. Participants who have better computer/internet knowledge are more satisfied with the smart glasses and AR applications. There is no significant interaction between all these factors. Since most participants have a moderate level or better than moderate level of knowledge using computers and the internet, it can be predicted that most educated people can easily accept smart glasses and AR applications.

7. Conclusions

This study was started by noting the scarcity of AR applications for hands-on training. As a first step toward incorporating the recorded teaching activities into learning procedures, the AR application was developed on AR glasses. In this work, the Questionnaire for Smart Glasses User Satisfaction (SGUS) and Questionnaire for User Interaction Satisfaction (QUIS) were investigated for augmented reality applications using Microsoft HoloLens.

The results of this study show that the approach is feasible. The experts wore the AR glasses to show the process, and the activities were recorded. The AR applications can facilitate the students to learn the process. The results show that the satisfaction of both teaching and learning are acceptable. The results indicate that satisfaction does increase when participants have higher computer knowledge levels. It also shows that gender, age, education level, and roles of students or experts do not have any effect on user satisfaction.

Author Contributions: H.X., P.S., F.W. contributed in the conceptualizing, writing and methodology. H.X. and F.W. performed the analysis. H.X. did validation and visualization. P.S. and F.W. helped in the review.

Funding: This study is part of the WEKIT (Wearable Experience for Knowledge Intensive Training) project. It is a European project supported under Horizon 2020 and is aiming to enable smart wearable technology for industrial training.

Acknowledgments: We express sincere appreciation to Kaj Helin, Principal Scientist, Project Manager of VTT Technical Research Centre of Finland, and Timo Kuula, Research Scientist of VTT Technical Research Centre of Finland for their modification of questionnaire QUIS and their design of questionnaire SGUS. We thank all the participants and companies for their valuable contributions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

331   The following abbreviations are used in this manuscript:

332

AR	Augmented Reality
UI	user interface
QUIS	Questionnaire for User Interaction Satisfaction
SGUS	Smart Glasses User Satisfaction
WEKIT	Wearable Experience for Knowledge Intensive Training

333

Appendix

334

Table A1. Spearman's rank coefficient of correlation of QUIS: the first value of each row represents Spearman's correlation coefficient, and the second value of each row represents the p value.

	QS1	QS2	QS3	QS4	QS5	QS6	QS7	QS8	QS9	QS10	QS11	QS12	QS13	QS14	QS15
QS1	1	0.39 0.000**	0.53 0.000**	0.50 0.000**	0.49 0.000**	0.34 0.000**	0.53 0.000**	0.47 0.000**	0.17 0.05	0.37 0.000**	0.37 0.000**	0.28 0.001	0.31 0.000**	0.32 0.000**	0.44 0.000**
QS2		1	0.53 0.000**	0.41 0.000**	0.34 0.000**	0.30 0.000**	0.37 0.000**	0.43 0.000**	0.16 0.07	0.35 0.000**	0.25 0.003	0.33 0.000**	0.58 0.000**	0.50 0.000**	0.52 0.000**
QS3			1	0.56 0.000**	0.55 0.000**	0.39 0.000**	0.49 0.000**	0.45 0.000**	0.16 0.06	0.33 0.000**	0.27 0.001	0.22 0.009	0.35 0.000**	0.37 0.000**	0.40 0.000**
QS4				1	0.49 0.000**	0.23 0.008	0.42 0.000**	0.41 0.000**	0.18 0.04	0.35 0.000**	0.38 0.000**	0.30 0.000**	0.27 0.001	0.27 0.001	0.40 0.000**
QS5					1	0.22 0.01	0.41 0.000**	0.45 0.000**	0.14 0.11	0.22 0.01	0.13 0.12	0.14 0.10	0.24 0.005	0.36 0.000**	0.34 0.000**
QS6						1	0.36 0.000**	0.26 0.002	0.26 0.001	0.18 0.03	0.25 0.003	0.09 0.28	0.28 0.000**	0.33 0.000**	0.33 0.000*
QS7							1	0.54 0.000**	0.17 0.05	0.38 0.000**	0.39 0.000**	0.28 0.001	0.24 0.004	0.35 0.000**	0.44 0.000**
QS8								1	0.23 0.006	0.40 0.000**	0.26 0.002	0.33 0.000**	0.26 0.002	0.40 0.000**	0.43 0.000**
QS9									1	0.35 0.000**	0.31 0.000**	0.31 0.000	0.19 0.024	0.32 0.000**	0.24 0.005
QS10										1	0.57 0.000**	0.45 0.000**	0.27 0.001	0.29 0.001	0.44 0.000**
QS11											1	0.43 0.000**	0.25 0.003	0.33 0.000**	0.38 0.000**
QS12												1	0.34 0.000**	0.30 0.000**	0.42 0.000**
QS13													1	0.57 0.000**	0.48 0.000**
QS14														1	0.47 0.000**
QS15															1

Signif. codes: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

References

1. Azuma, Ronald T. "A survey of augmented reality." In *Teleoperators & Virtual Environments* 6.4 (1997): 355-385.
2. Sherstyuk A, Vincent D, Berg B, Treskunov A. Mixed reality manikins for Medical Education. In *Handbook of Augmented reality*; Springer: New York, NY, 2011; pp. 479-500, ISBN.
3. Ong SK, Shen Y, Zhang J, Nee AYC. Augmented Reality in Assistive Technology and Rehabilitation Engineering. In *Handbook of Augmented Reality*; Springer: New York, NY, 2011; pp. 603-630, ISBN.
4. Nilsson J, Ödblom ACE, Fredriksson J, Zafar A. Using Augmentation Techniques for Performance Evaluation in Automotive Safety. In *Handbook of Augmented Reality*; Springer: New York, NY, 2011; pp. 631-649, ISBN.
5. Ras E, Wild F, Stahl C, Baudet A. Bridging the Skills Gap of Workers in Industry 4.0 by Human Performance Augmentation Tools. *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '17* **2017**, doi:10.1145/3056540.3076192.
6. Perey C, Wild F, Helin K, Janak M, Davies P, Ryan P. Advanced manufacturing with augmented reality. *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* **2014**, doi:10.1109/ismar.2014.6948518.
7. Olsson T. Concepts and Subjective Measures for Evaluating User Experience of Mobile Augmented Reality Services. In *Human Factors in Augmented Reality Environments*; Springer: New York, NY, 2013; pp. 203-232, ISBN.
8. Augmented Reality: What Does It Mean for UX? In: Nielsen Norman Group. Available online: https://www.nngroup.com/articles/augmented-reality-ux/ (accessed on 5 Jul 2018).
9. Microsoft. Microsoft HoloLens. In: Microsoft HoloLens. Available online: https://www.microsoft.com/en-us/hololens/commercial-overview (accessed on 5 Jul 2018).
10. Rauschnabel, Philipp A.; Brem, Alexander; Ro, Young K. Augmented Reality Smart Glasses: Definition, Conceptual Insights, and Managerial Importance.Unpublished Working Paper, The University of Michigan-Dearborn, College of Business (2015). Available online: https://www.researchgate.net/profile/Alexander_Brem/publication/279942768_Augmented_Reality_Smart_Glasses_Definition_Conceptual_Insights_and_Managerial_Importance/links/5721ec2e08aee857c3b5dd6c/Augmented-Reality-Smart-Glasses-Definition-Conceptual-Insights-and-Managerial-Importance.pdf (accessed on 12 Jul 2018)
11. Rauschnabel PA, Brem A, Ivens BS. Who will buy smart glasses? Empirical results of two pre-market-entry studies on the role of personality in individual awareness and intended adoption of Google Glass wearables. *Computers in Human Behavior* **2015**, *49*, 635-647, https://doi.org/10.1016/j.chb.2015.03.003
12. Wiederhold, Brenda K. "Time to port augmented reality health apps to smart glasses?." *Cyberpsychology, Behavior, and Social Networking* **2013**, *Volume: 16 Issue 3*, 157-158, http://doi.org/10.1089/cyber.2013.1503
13. Questionnaire For User Interaction Satisfaction. Available online: http://lap.umd.edu/quis/ (accessed on 29 Jan 2018).
14. Ssemugabi S, de Villiers R. A comparative study of two usability evaluation methods using a web-based e-learning application. *Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries - SAICSIT '07* **2007**, doi:10.1145/1292491.1292507.
15. Fominykh M. D2.4 First Prototype. Available online: http://wekit.eu/d2-4first-prototype/ (accessed on 2 Sep 2018).
16. WEKIT D1.4 Requirements for Scenarios and Prototypes. Available online: http://wekit.studiohangloose.it/wp-content/uploads/2017/06/WEKIT_D1.4.pdf (accessed on 7 Nov 2017).
17. WEKIT D6.4 Implementation of Evaluation Trials in Aeronautics. Available online: http://wekit.eu/wp-content/uploads/2017/09/WEKIT_D6.4 (accessed on 7 Nov 2017).
18. WEKIT D6.2 Annex 1 Training scenario and Evaluation Plan for Engineering. Available online: http://wekit.studiohangloose.it/wp-content/uploads/2017/06/WEKIT_D6.2.pdf (accessed on 9 Nov 2017).
19. WEKIT D6.5 Implementation of Evaluation Trials in Engineering. Available online: http://wekit.eu/wp-content/uploads/2017/09/WEKIT_D6.5 (accessed on 9 Nov 2017).
20. WEKIT D6.6 Implementation of Evaluation Trials in Space. Available online: http://wekit.eu/wp-content/uploads/2017/09/WEKIT_D6.6 (accessed on 5 Jun 2018).
21. Chin, John P., Virginia A. Diehl, and Kent L. Norman. In *Proceedings of the SIGCHI conference on Human factors in computing systems*; ACM: New York, USA, 1988; pp. 213-218, ISBN:0-201-14237-6.