D6.12 Evaluation
Process and Results for
Space case

Editors: Liliana Ravagnolo
## Revision History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Contributor(s)</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>15.02.2019</td>
<td>Liliana Ravagnolo (AL)</td>
<td>Draft version including contributions for SUS data and interviews by Timo Kuula (VTT)</td>
</tr>
<tr>
<td>0.3</td>
<td>01.03.2019</td>
<td>Liliana Ravagnolo (AL)</td>
<td>Including preliminary comments from GFT</td>
</tr>
<tr>
<td>0.4</td>
<td>09.03.2019</td>
<td>Alla Vovk, Fridolin Wild, Will Guest (OBU)</td>
<td>Including data analysis (OBU)</td>
</tr>
<tr>
<td>0.5</td>
<td>10.03.2019</td>
<td>Bibeg Limbu, Roland Klemke (OUNL)</td>
<td>Including data analysis (OUNL)</td>
</tr>
<tr>
<td>0.7</td>
<td>11.03.2019</td>
<td>Mark Ransley, Jazz Razool, Carl Smith (RAV)</td>
<td>Including sensors data analysis (RAV)</td>
</tr>
<tr>
<td>0.8</td>
<td>13.03.2019</td>
<td>Liliana Ravagnolo (AL)</td>
<td>Including contributions from VTT/OBU/OUNL/RAV</td>
</tr>
<tr>
<td>0.9</td>
<td>14.03.2019</td>
<td>Liliana Ravagnolo (AL)</td>
<td>Including comments from peer review</td>
</tr>
<tr>
<td>1.0</td>
<td>13.03.2019</td>
<td>Mikhail Fominykh (EP)</td>
<td>Quality review, styles and formatting</td>
</tr>
</tbody>
</table>

Disclaimer: All information included in this document is subject to change without notice. The Members of the WEKIT Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the WEKIT Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.
D6.12 Evaluation Process and Results for Space case

WP 6 | D6.12

Editors:
Liliana Ravagnolo (AL), Timo Kuula (VTT), Alla Vovk (OBU), Bibeg Limbu (OUNL),
Mark Ransley (RAV)

Authors:
Liliana Ravagnolo (AL), Timo Kuula (VTT), Alla Vovk (OBU), Fridolin Wild (OBU), Will Guest
(OBU), Bibeg Limbu (OUNL), Roland Klemke (OUNL), Mark Ransley (RAV), Jazz Razool (RAV), Carl
Smith (Rav), Mikhail Fominykh (EP)

Reviewers:
Maurizio Megliola (GFT)
Paul Lefrere (CCA)

<table>
<thead>
<tr>
<th>Deliverable number</th>
<th>D6.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissemination level</td>
<td>Public</td>
</tr>
<tr>
<td>Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Status</td>
<td>final</td>
</tr>
<tr>
<td>Date</td>
<td>14.03.2019</td>
</tr>
<tr>
<td>Due date</td>
<td>M39</td>
</tr>
</tbody>
</table>
Contents

REVISION HISTORY .......................................................... 2
EXECUTIVE SUMMARY ..................................................... 5

1. TRIAL DESCRIPTION ....................................................... 6
   1.1. USE CASE DESCRIPTION ........................................... 6
   1.2. SPACE PROCEDURE DESCRIPTION ............................... 8
   1.3. SPACE PROCEDURE DESCRIPTION (CONTROL GROUP) ....... 11
   1.4. TEST POPULATION ................................................ 15
   1.5. TEST DESCRIPTION ................................................. 16
       1.5.1. Trial Set-up .................................................. 16
       1.5.2. Trial execution .............................................. 17
       1.5.3. Prolonged Hololens testing ............................... 19

2. DATA COLLECTION AND EVALUATION .................................. 20
   2.1. WEKIT QUESTIONNAIRES ......................................... 20
   2.2. LIME SURVEY TOOL ............................................... 21
   2.3. TRAINING QUESTIONNAIRE ...................................... 21

3. RESULTS ........................................................................ 24
   3.1. QUANTITATIVE RESULTS: QUESTIONNAIRES ................. 25
       3.1.1. Technology Acceptance Model for AR/WT (TAMARA) .... 25
       3.1.2. System Usability Scale (SUS) ............................... 26
       3.1.3. Spatial Interface Evaluation (SPINE) ....................... 27
       3.1.4. Knowledge Assessment Results ............................ 30
       3.1.5. Expert Evaluation ............................................ 31
   3.2. INTERVIEWS RESULTS ............................................. 33
   3.3. BIOSENSOR DATA .................................................. 35

4. DISCUSSION ................................................................... 35
   4.1. QUANTITATIVE RESULTS: QUESTIONNAIRES ............... 35
       4.1.1. Technology Acceptance Model for AR/WT (TAMARA) .... 35
       4.1.2. System Usability Scale ....................................... 36
       4.1.3. Spatial Interface Evaluation (SPINE) ....................... 36
       4.1.4. Knowledge Assessment Results ............................ 37
       4.1.5. Expert Evaluation ............................................ 38
   4.2. INTERVIEWS ......................................................... 38
   4.3. BIOSENSOR .......................................................... 39

5. CONCLUSIONS .............................................................. 39

REFERENCES .................................................................... 41
Executive summary

The purpose of this document is to describe the second and final trial of the WEKIT prototype for the Space Industrial Case. The selected use case for the test (including the after-action review) covered all features of the WEKIT platform and learning model. It was a procedure derived from the future manned Exploration activities, dedicated to the recharge of the batteries of a Mars Rover before starting an exploration mission on the planet. The users had to execute the steps of the procedure using the Player functionality, while trainers performed the creation of the steps using the Recorder functionality. The scenario and the procedure were already described in D6.9, although they have been updated and adapted to better represent the scenario.

The challenge for each trial was:

a) to explore the use of AR/WT to achieve the necessary level of trainee performance, in less time and making fewer errors than conventional training;

b) to help trainees in using the Wearable Experience methodology to record critical events and share them with their trainer and other trainees;

c) to extensively experiment with the use of the WEKIT recorder by experts to create authoritative sets of training procedures.

All aspects of that challenge were met in each of the trials as will be presented in the following deliverable. The space and aero trials had in common the use of the player to show the performance standards that trainees had to meet and the expected measurements associated with each work flow (as in the ‘after-action reviews’ that are required in many safety-critical contexts and in field-service maintenance). Of course, while standardisation of workflows and training is key in health engineering too, there can be much wider variation in the presenting conditions of hospital patients, adding complexity to the patient-side of the health trial.

The Space Industrial Case trial took place at ALTEC facility in Turin, during an extensive period, starting from July 4\textsuperscript{th} 2018 and ending on February 6\textsuperscript{th} 2019. This longer timeframe was needed in order to comply with the requirement of having 150 candidates for each industrial partner to sustain the test. In ALTEC we had a total amount of 199 candidates involved in the trials, distributed as follows:

- 147 candidates tested the Player application. Among them 30 candidates were coming from ALTEC and Thales Alenia Space employees, all the others were students from Politecnico of Turin or people interested in space activities contacted through social forums and platforms.
- 30 candidates were part of the Control Group.
- 22 candidates (all ALTEC internal experts) tested the Recorder application.

Since September 2018 the candidates used also the sensors harness during the performance, so that the relevant sensors data have been collected and analysed as well.

The data and feedback from the participants were collected by means of online surveys designed in order to get a more comprehensive feedback on their experience. The data collected from the participants have been later on analysed and the results are described in this report.
1. Trial description

1.1. Use case description

The objective of this trial for the space pilot case is to test the final WEKIT prototype on the space scenario defined in D6.9. The chosen scenario is considered a valid example of a potential astronaut training activity to be performed on the Mars surface during an exploration mission.

The activity analysed and detailed for the WEKIT scenario is a procedure that foresees the recharge of the batteries of a Mars Rover prior to starting an exploration mission on the planet. Before performing the activity, some safety checks are needed in order to put the Rover in safe state. After the recharge is completed some additional steps are needed in order to put the Rover back in operational state.

All the components and mock-ups are available at ALTEC facility as well as the complete procedure, which is performed on the ALTEC Mars Moon Terrain Demonstrator (MMTD), an arena designed on purpose to simulate the Martian soil, used normally to test the locomotion of Mars Rovers and plan their activities. This kind of procedure allows taking into account several aspects of the training methodology and also testing the different features of the WEKIT prototype.

![Figure 1. ALTEC Mars Moon Terrain Demonstrator (MMTD)](image)

The procedure describes in detail the actions that have to be executed for capturing the trainer experience and for re-enacting the trainer experience (trainee) with the WEKIT prototype. From the user's point of view, the WEKIT.one software prototype comprises of two main functionalities: the Recorder, that allows the trainer to create the procedure, and the Player, that allows the trainee to execute the procedure following the instructions recorded by the trainer.
In addition, a set of sensors have been identified and used during the procedure in order to collect biometrical data. As described in D4.4 “The WEKIT prototype garment, designed and prototyped by the WEKIT consortium, houses sensors to gather data that can be used to determine psychological and physical metrics that contribute to identifying a learner’s state. Specifically, these metrics are Temperature, Heart Rate and Galvanic Skin Response (GSR), with additional IMU sensors for determining posture and correcting when the previous sensor readings are corrupted by motion artefacts. Trial data are collected using three categories of sensors. These categories are human, environment and device sensors. Integrated into the garment are sensors to detect heart activity, Galvanic Skin Response, electro-myographic activity for the forearm and hand, positional tracking, 9-axis inertial movement as well as temperature and humidity. The sensor ecosystem has been integrated with the design of the wearable prototype garment”.

Due to the late delivery of the Recorder application, that was made available to the team on 31st October 2018, the ALTEC trials have been initially organized using the Player functionality, based on the Mars Rover procedure depicted in the following paragraph. After the Recorder was made available, a new learning activity version was prepared by the ALTEC instructors utilizing the recorder functionality.
1.2. Space procedure description

1.2.1. INTRO

- The Mars rover is starting an exploration mission. YOUR TASK: Evaluate the status of the solar panels and charge the battery.
- In the picture of the rover, you will see the location of the important points related to your task. Identify coordinate reference system (front, rear, left, right based on rover cockpit)
- You can now look through all the steps of the task. You will be guided while performing the task.

1.2.2. Step 1/15

- Walk to the Mars rover. Watch your step while performing the task.
- Show the path to reach the rover avoiding obstacles (rocks).

1.2.3. Step 2/15

- Visual guidance: Localize the control panel on the rear side of the rover
- Find the control panel. Activate rover brakes removing the red key from the control panel.
- Record instructor hand-arm position and movement through dedicated sensor. Compare with trainee performance.

1.2.4. Step 3/15

- Visual guidance: Localize the emergency/safety red button on the rover rear right side)
- Press the emergency/safety red button located on the rover rear right side to force the rover in OFF status.
- Record instructor hand-arm position and movement through dedicated sensor. Compare with trainee performance. Press the button to lock.
1.2.5. Step 4/15
- Visually inspect the solar panels right side and verify that they are undamaged (no scratches, holes).
- Visually inspect the solar panels left side and verify that they are undamaged (no scratches, holes).
- *Show pictures of damaged solar panels for comparison.*

1.2.6. Step 5/15
- Retrieve the battery charger device box located on the front side of the rover.
- *Highlight path and box location*

1.2.7. Step 6/15
- Identify the rover red battery charge connector on the left side of the rover
- *Highlight path and red battery charger connector position on the rover (arrow pointing position).*

1.2.8. Step 7/15
- Unlock the box and extract the battery charger.
- *Show video of opening box and battery charger extraction.*

1.2.9. Step 8/15
- Place the battery charger on the wheel close to the rover battery charger connector.
- *Show video highlighting correct battery charger positioning on wheel.*

1.2.10. Step 9/15
- Connect the battery charger to the rover red battery charge connector.
- *Record instructor hand-arm position and movement through dedicated sensor. Compare with trainee performance.*

1.2.11. Step 10/15
- Press the button MODE on the battery charger to start the charging process.
- *Text: Charging process started.*

1.2.12. Step 11/15
- Text: charging time is completed.
- *Press the button MODE on the battery charger to complete the charging process.*
1.2.13. Step 12/15

- Disconnect the battery charger from the red battery charge connector.
- Record instructor hand-arm position and movement through dedicated sensor. Compare with trainee performance.

1.2.14. Step 13/15

- Store the battery charger in the box.
- Show video of battery charger insertion in the box and box closure.

1.2.15. Step 14/15

- Visual guidance: Localize the emergency/safety red buttons on the rover rear side (left and right side of the rover)
- Release the emergency/safety red button located on the rover rear right side to enable the rover operative status.
- Show video of push and turn clockwise the button to release.
- Record instructor hand-arm position and movement through dedicated sensor. Compare with trainee performance.

1.2.16. Step 15/15

- Visual guidance: Localize the control panel on the rear side of the rover
- Find the control panel. Release rover brakes inserting the red key in the control panel.
- Record instructor hand-arm position and movement through dedicated sensor. Compare with trainee performance.

*TASK COMPLETED*

In addition to the nominal procedure described above, VTT inserted the possibility of raising an alarm at any time (randomly decided by the instructor, activating a dedicated application running on a cellular phone). The alarm was shown on the Hololens as a message derived from temperature sensors going out of scale. When the alarm was raised the application was stopping the nominal procedure flow, automatically activating the “contingency” procedure that required the candidate to use an external box (called “Safety box”) that the candidate was requested to bring with him/her during the whole procedure. The software guided the candidate on how to perform a reset of the safety box. If the reset was performed correctly, this was detected via the sensor communication, bringing back the sensor data to nominal. After that, the activity would resume back to the main procedure, at the same step where the alarm was raised, to continue with the activity that was interrupted.

The sub-routine maintenance procedure was described as follows:

**External event**

**Warning message: Suit temperature below nominal. Activate maintenance procedure.**

Interrupt nominal procedure and start maintenance procedure.
Perform maintenance tasks on “safety box”.

**Warning message: Suit temperature nominal. Return to nominal tasks.**

### 1.3. Space procedure description (Control Group)

In order to verify the effectiveness of the training performed using the AR tool, a parallel procedure was developed to be used by a selected control group. 30 participants, randomly selected, were asked to complete the procedure following instructions from a paper version. After having completed the procedure they were asked to fill the dedicated on-line survey (see also Section 2 for additional details) mainly dedicated to analyse the training retention. Only after that, they had the opportunity to test out the WEKIT Player application, thus avoiding contamination of their ratings.

The procedure prepared for the Control Group was the following:

#### 1.3.1. INTRO

- The Mars rover is starting an exploration mission. YOUR TASK: Evaluate the status of the solar panels and charge the battery.
- In the picture of the rover, you can see the location of the important points related to your task. Identify coordinate reference system (front, rear, left, right based on rover cockpit)

- You can now look through all the steps of the task. You will be guided while performing the task.

#### 1.3.2. Step 1/15

- Walk to the Mars rover. Watch your step while performing the task.

#### 1.3.3. Step 2/15

- Find the control panel. Activate rover brakes removing the red key from the control panel.
1.3.4. Step 3/15

- Press the emergency/safety red button located on the rover rear right side to force the rover in OFF status.

1.3.5. Step 4/15

- Visually inspect the solar panels right and left side and verify that they are undamaged (no scratches, holes).
1.3.6. Step 5/15
- Retrieve the battery charger device box located on the front side of the rover.

1.3.7. Step 6/15
- Identify the rover red battery charge connector on the left side of the rover

1.3.8. Step 7/15
- Unlock the box and extract the battery charger.

1.3.9. Step 8/15
- Place the battery charger on the wheel close to the rover battery charger connector.
1.3.10. Step 9/15

♦ Connect the battery charger to the rover red battery charge connector.

1.3.11. Step 10/15

♦ Press the button MODE on the battery charger to start the charging process. Charging time of 2 hrs. is simulated.

1.3.12. Step 11/15

♦ Press the button MODE on the battery charger to complete the charging process.

1.3.13. Step 12/15

♦ Disconnect the battery charger from the red battery charge connector.

1.3.14. Step 13/15

♦ Store the battery charger in the box.

1.3.15. Step 14/15

♦ Release the emergency/safety red button located on the rover rear right side to enable the rover operative status.
1.3.16. Step 15/15

♦ Find the control panel. Release rover brakes inserting the red key in the control panel.

*TASK COMPLETED*

In this case no alarm was raised during the execution and no maintenance procedure was required.

1.4. Test population

The test population identified for this second trial of the WEKIT prototype came from different groups: experts, students, and “average people” with no specific background in space activities, beside a generic interest in space activities. In fact, the goal of this second trial was to demonstrate that the WEKIT tool can provide just-in-time effective and efficient training also to people who have no specific background and preparation related to the activity to be performed. Just following the instructions displayed on the Player application, an average (“normal”) candidate can perform correctly a complex task without problems.

The “expert” definition refers to all the ALTEC and Thales Alenia Space Italy (TAS-I) employees that were involved in the test. 30 candidates tested the Player application in July 2018 and 22 tested the Recorder application in January 2019. Expert candidates were selected because of their aero spatial background or because they were directly involved in space programs and studies, where the technology that we are developing within the WEKIT project might be useful. Other experts were chosen because they had a computer science background and in some cases they also had knowledge of AR and VR technologies.

The selected students were coming from University (mainly Politecnico of Turin) specialization programs post high school graduation and high school programs. In addition a large number of people generically interested in space activities were recruited through social media and web forums, in order to reach the required number of 150 participants per industrial scenario. In total the participants belonging to those two groups were 147 (117 tested the Player application and 30 were part of the Control Group).

The team conducting the trials from July 2018 up to February 6th 2019 was composed by skilled astronaut trainers, so their role in the WEKIT trial was on one side that of a real expert for this kind of activities and scenario, and on the other side to support the candidates who had no experience of this type of training.
1.5. Test description

1.5.1. Trial Set-up

The trial was performed according to the General Data Protection Regulation (EU) 2016/679 ("GDPR") which was adopted by the EEC on 14 April 2016, and became enforceable beginning 25 May 2018. The rules and regulations about data privacy management were followed. In fact, before the test, the following preliminary actions were done in order to be compliant to the present law:

- Formal assignment from the company of the WEKIT data processing responsibilities to Dr. Liliana Ravagnolo, responsible of the trials organization and conduction in ALTEC
- Preparation of a dedicated "Informed Consent Form" containing detailed information related to the WEKIT project, the scope of the trials, the data acquisition and the use of images. The consent form included specific permission to the use of images and sensors data (although totally anonymous) to be signed by each participants before performing the test.
- Preparation of a complete "Disclosure of risks related to the Martian scenario" form, prepared by the ALTEC Responsible for the Prevention and Protection Service (RSPP) describing in details all potential risks related to the test. Each candidate was required to sign for acceptance the form before starting the test.

ALTEC second trial period took place in Turin starting from July 2nd 2018. During the first week the initial two days were used for setting up the prototype, testing the Player application and preparing the working area. Three additional days were used for the real test with ALTEC and TAS-I employees.

Six members of the WEKIT Consortium took part in the trials start-up in Turin and everyone had an active role during the setup and test execution. Other consortium members joined in different...
periods during the trials execution, in early August and early September timeframes, mainly to test the Recorder application usability and the garment sensors data acquisition.

![Figure 4. Trial execution](image)

### 1.5.2. Trial execution

The WEKIT.one software prototype comprises of two key functionalities: the WEKIT recorder (used by the trainer to record the procedure) and the WEKIT player (used by the trainee to wear the expert experience). For more information about the two applications, see D2.5 - Final Prototype.

The trial took place at ALTEC facility in Turin, during an extensive period, starting from July 4th 2018 and ending on February 6th 2019. This longer timeframe was needed in order to comply with the requirement of having 150 candidates for each industrial partner to sustain the test. In ALTEC we had a total amount of 199 candidates involved in the trials, distributed as follows:

- **147 candidates tested the Player application.** Among them 30 candidates were coming from ALTEC and Thales Alenia Space employees, all the others were students from Politecnico of Turin or people interested in space activities contacted through social forums and platforms.
- **30 candidates were part of the Control Group.**
- **22 candidates (all ALTEC internal experts) tested the Recorder application.**

As mentioned in Section 1.1, due to the late delivery of the Recorder application, that was made available to the team on 31st October 2018, the ALTEC trials started in July 2018 using an earlier player version, based on the Mars Rover procedure depicted in Section 1.2. Each trial session was organized as follows:

**PLAYER TESTING:**

1. Introduce participant to the project (purpose/data storage/introduce Hololens and wearable tech/signature of consent form) (duration 5-10 minutes)
2. Perform gesture training to familiarize with the Hololens usage (duration 5-10 minutes)
3. Perform space procedure using the Player application (duration 15-20 minutes)
4. On line Survey filling (Questionnaire – duration 10-15 minutes)

CONTROL GROUP:

A randomly selected group of participants (30 in total) were assigned to the Control Group. In this case they were required to perform the activity utilizing the paper procedure described in para 1.3 then they had to complete the dedicated questionnaire that was mainly devoted to collect information about the training retention (evaluating what they could remind/have learned while performing the task). After that they were asked to test the WEKIT tool using the player application, providing verbal feedbacks comparing the performance (and understanding) of the procedure conducted on paper procedure and using the Player application. The whole activity was organized as follows:

1. Introduce participant to the project (purpose/data storage/introduce Hololens and wearable tech/signature of consent form) (duration 5-10 minutes)
2. Perform space procedure using the paper procedure (duration 15-20 minutes)
3. On line Survey filling (Questionnaire – duration 10-15 minutes)
4. Perform gesture training to familiarize with the Hololens usage (duration 5-10 minutes)
5. Perform space procedure using the Player application (duration 15-20 minutes)

Concerning the garment testing, during the first trial week in July 2018 we tested the initial design consisting of a tight-fitting interior layer of thin Lycra to house the skin-contact sensors and a vest on the outside to house the bulkier components. (Please refer to deliverable 5.8 for additional details).

However, the tests demonstrated that the use of complete garments presented numerous challenges, specifically providing inadequate means of securing sensors to a wide range of body shapes over wide-ranging environmental conditions, and additionally we found each design iteration to be costly and time-consuming.

Thus, a final garment design was produced and made available for testing in ALTEC in September 2018. This version is based on an adjustable harness, which was found to be far cheaper, more easily manufactured and more robust to variations in body shape and environmental temperature. So from September on also the sensor data acquisition started and continued up to the end of the trial period (February 2019).
As soon as the final version of the Recorder application was available, the ALTEC team started testing the application inviting only ALTEC internal experts with extensive experience in space operations and procedures definition. Most of them had already tested the Player application in July 2018, others approached for the first time the WEKIT environment. In this second case we decided to provide them with the whole experience, to make them aware of the goal of the activity (to create a new Player version similar for completeness and quality to the one used during the Player testing).

**RECORDER TESTING:**

1. Introduce participant to the scope of the Recorder
2. If needed, refresh gesture training
3. Explain how the menu works, how to create task stations and how to add various annotations, provide paper procedure to be followed.
4. Trial starts, record as many task stations as possible in the available time (normally around 1 hour per candidate).
5. After the trial ends, show the result of the recording with the Player application
6. On line Survey filling (Questionnaire – duration 10-15 minutes)

### 1.5.3. Prolonged Hololens testing

Beside the results related to the usage of the WEKIT tool, that will be described in the following section, the long trial period in ALTEC provided the possibility to test the Hololens reaction under very intensive usage and in very different ambient conditions.

The Mars Terrain Simulator in ALTEC is a huge facility (the area exact dimensions are 945 m² and 33,200 m³), which is normally not air-conditioned due to the very high costs related. This second trial session was conducted both in summer (with very high temperatures inside, reaching around 38°) and in wintertime (were the internal temperature was very low, sometimes around 4-5°).
This situation allowed indirectly the possibility to test the Hololens reactions to such extreme conditions (out of the manufacturer's environmental range).

In general, we demonstrated that the Hololens starts having problems when it is forced to work in extreme temperature conditions. When the temperature is too high, device starts overheating and after few hours of usage it became unresponsive up to a complete automatic shutdown.

On the other side of the scale, when the temperature is extremely low, the battery drains very rapidly, especially when using the Recorder application, that is much more demanding in terms of memory and CPU usage. To remediate this situation, we used them connected to battery packs to allow the testing of up to six candidates per day.

2. Data collection and evaluation

2.1. WEKIT Questionnaires

During the first trial phase paper questionnaires were used to collect feedback from the participants and to enable the evaluation process after the trial. The questionnaire comprises of seven sections, derived from six different questionnaires:

- **SSQ**: the Simulator Sickness Questionnaire is used to check the symptoms, which are commonly experienced by users of virtual and augmented reality systems. The SSQ questions were asked before the trial and after the trial to check if the user is healthy (pre) and then to evaluate any possible symptoms (post).
- **TAMARA**: the Technology Acceptance Model for AR/WT questionnaire contains some questions to check the technology acceptance of the user toward the proposed prototype. The questionnaire has been developed in Wild et al. (2017).
- **SUS**: the System Usability Scale questionnaire contains some questions to measure the usability of the proposed prototype.
- **QUIS**: the Questionnaire for User Interaction Satisfaction assesses users’ subjective satisfaction with specific aspects of the human-computer interface
- **TM**: the Transfer Mechanisms questionnaire was designed to evaluate the implementation of the prototype according to the WEKIT framework. The questionnaire consisted of statements stating the optimal implementation of each Transfer mechanism.

In some case the users were also interviewed by the WEKIT team, after filling the questionnaire, in order to get a more detailed feedback on the experience. The SSQ questionnaire was filled twice (before and after the test) to collect feedbacks about possible simulator sickness caused by the Hololens usage.
2.2. Lime Survey Tool

Before starting the trial two phase, the Consortium decided that, due to the high expected number of participants (450 in total, 150 for each industrial partner), in order to facilitate the analysis of the results and elaborate automatically the questionnaires, it was more convenient to realize an on-line survey containing all the relevant questionnaires to be filled by the candidates.

The platform utilized was the Lime Survey Tool and an update/revision of the questionnaires used during trial one was done. The survey used for trial two is composed by six sections:

- Demographics data
- Learning Model Evaluation
- TAMARA – Technology Acceptance Model for AR/WT
- SPINE - Spatial User Interface of the Augmented Reality application
- SUS – System Usability Scale
- Training Questionnaires (one for each industrial case scenario) – To evaluate training effectiveness

The Simulator Sickness Questionnaire (SSQ) was removed since the first trial indicated that there are no significant levels of simulator sickness the participants suffer of after using the Hololens applications.

2.3. Training Questionnaire

A dedicated Training Questionnaire was developed by each industrial partner with the scope of assessing the training retention after the session. The questionnaire contained several questions related to the procedure to be performed and was filled by all candidates. The expectation is that the results to the training questionnaire should be better for the candidates who have performed the task with the use of the Player application (compared to the paper procedure, used by the Control Group). The assumption is that the additional material (photo, video, audio information) provided with the AR tool should improve the learning curve related to the procedure activities.

The following questions were prepared by the ALTEC instructors to check the training retention related to the space scenario:
MARS SCENARIO TRAINING RETENTION QUESTIONNAIRE

1. On the rover picture provided, please identify the correct rover coordinate system.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Rover Coordinate System A" /></td>
<td><img src="image2" alt="Rover Coordinate System B" /></td>
<td><img src="image3" alt="Rover Coordinate System C" /></td>
</tr>
</tbody>
</table>

2. On the rover control panel picture provided, please identify the correct location of the red key.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4" alt="Rover Control Panel A" /></td>
<td><img src="image5" alt="Rover Control Panel B" /></td>
<td><img src="image6" alt="Rover Control Panel C" /></td>
</tr>
</tbody>
</table>

3. On the rover picture provided, please locate the red battery charger connector.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image7" alt="Rover Battery Charger Connector A" /></td>
<td><img src="image8" alt="Rover Battery Charger Connector B" /></td>
<td><img src="image9" alt="Rover Battery Charger Connector C" /></td>
</tr>
</tbody>
</table>
4. On the rover picture provided, please locate the red safety buttons.

On the rover picture provided, please locate the red safety buttons.

A  B  C

5. What happens when you remove the red key? (switch off the rover, activate the brakes, release the brakes)
6. What should you do to force the rover in OFF status? (Remove Red key, Press red button rear left side, press red button rear right side, press both red buttons on rear left and right side).
7. What should you visually inspect before starting the procedure? (battery, solar panels, battery charger)
8. Identify possible damages that could affect rover solar panels on Mars. (object impact-rip, sand-rip, object impact-dust)
9. Which colour is the rover battery charge connector? (blue, red, yellow,)
10. Where do you position the rover battery charger? (Close to control panel, close to rear side, close to the wheel)
11. After connecting the battery charger to the battery charge connector the operator should... (press the MODE button on the battery charger, press the MODE button on the control panel, wait for information).
12. Which warning was triggered during the procedure? (high temperature, high pressure, low temp, low press)
13. What did you do after alarm? (continued recharging procedure, asked for instruction, executed the maintenance procedure).
14. How would you move to the following procedure step? (voice command next, mouse clicking, wait for timeout)
15. How will you reactivate the rover at recharging procedure completed? (push and turn red safety button clockwise, push and turn red safety button counter-clockwise, turn and pull red safety button clockwise, turn and pull red safety button counter-clockwise).

A dedicated correlation table was created to assess the correspondence among the questions and the procedure step.

Table 1. Correlation Table

<table>
<thead>
<tr>
<th>Assessment question</th>
<th>Procedure step</th>
<th>Correct answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. On the rover picture provided, please identify the rover coordinate system.</td>
<td>Intro</td>
<td>A</td>
</tr>
</tbody>
</table>
2. On the rover picture provided, please identify the correct location of the red key.  
3. On the rover picture provided, please locate the red battery charge connector.  
4. On the rover picture provided, please locate the red safety buttons.  
5. What happens when you remove the red key?  
6. What should you do to force the rover in OFF status?  
7. What should you visually inspect before starting the procedure?  
8. Identify possible damages that could affect rover solar panels on Mars.  
9. Which colour is the rover battery charge connector?  
10. Where do you position the rover battery charger?  
11. After connecting the battery charger to the battery charge connector the operator should...  
12. Which warning was triggered during the procedure?  
13. What did you do after alarm?  
14. How would you move to the following procedure step?  
15. How will you reactivate the rover at recharging procedure completed?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Step 2/15</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>Step 6/15</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>Step 3/15</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>Step 2/15</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>Step 3/15</td>
<td>B</td>
</tr>
<tr>
<td>7</td>
<td>Step 4/15</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>Step 4/15</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>Step 6/15</td>
<td>B</td>
</tr>
<tr>
<td>10</td>
<td>Step 8/15</td>
<td>C</td>
</tr>
<tr>
<td>11</td>
<td>Step 10/15</td>
<td>A</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>A</td>
</tr>
</tbody>
</table>

### 3. Results

In total **199 data entries** were considered for analysis for the AR condition group.

**Gender.** In total among **199** data entries we had **42** females and **156** males. 1 person did not want to declare their gender.

**Age.** “18-24” - **64** subjects, “25-34” - **50** subjects, “35-44” - **33** subjects, “45-54” - **32** subjects, “55-64” - **19** subjects, “65+” - **1** subject.

**Education.** Primary - **1** subject, Upper secondary degree - **67** subjects, Bachelor degree - **51** subjects, Master degree - **64** subjects, PhD degree - **14** subjects, other - **2** subjects.

**Experience with AR.** We asked participants about their experience with augmented reality hardware. In total **22 subjects** never tried AR head-mounted display (HMD) before, and **29 subjects** indicated they had a chance to try AR HMD, **148** subjects left the question blank.

**Computer knowledge.** From those, who completed the question about their general computer skills, **4** participants described their computer knowledge as “poor”, **36** reported it as “moderate”, **93** as “good”, and **65** as “very good”.

**Type of AR user.** We asked participants to identify their experience with AR and how they can describe themselves as AR user. In total, **22** participants described themselves as “Novice user”, **19** as “Casual users”, and **2** participants as “Frequent user”, **8** “Expert users”. 148 subjects left the question blank.
3.1. Quantitative Results: Questionnaires

3.1.1. Technology Acceptance Model for AR/WT (TAMARA)

The Technology Acceptance Model for AR/WT (TAMARA) is a tool for measuring the generic technology acceptance models (see Wild et al., 2017).

All items except one use a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7), while one item, usage frequency, is rated on a 6-point scale. The item "usage frequency" (UF1), allows expressing in what intervals the participants turn to using AR and WT. They are presented to the experiment participants in the questionnaire in the following manner:

**ATU4** I look forward to those aspects of my job that require me to use AR & WT.

<table>
<thead>
<tr>
<th>strongly disagree (1)</th>
<th>disagree (2)</th>
<th>somewhat disagree (3)</th>
<th>neither agree or disagree (4)</th>
<th>somewhat agree (5)</th>
<th>Agree (6)</th>
<th>Strongly agree (7)</th>
</tr>
</thead>
</table>

The full item set (with corresponding codes) are listed below:

<table>
<thead>
<tr>
<th>CODE</th>
<th>STATEMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATU4</td>
<td>I look forward to those aspects of my job that require me to use AR &amp; WT.</td>
</tr>
<tr>
<td>CSE4</td>
<td>I could complete a job, if I had used similar technologies before this one to do the same job.</td>
</tr>
<tr>
<td>EE2</td>
<td>My interaction with AR &amp; WT is clear and understandable.</td>
</tr>
<tr>
<td>FC1</td>
<td>I have the resources necessary to use AR &amp; WT.</td>
</tr>
<tr>
<td>HM2b</td>
<td>I like working with AR &amp; WT.</td>
</tr>
<tr>
<td>HT2</td>
<td>I am addicted to using AR &amp; WT.</td>
</tr>
<tr>
<td>IMG1</td>
<td>People in my organization who use AR &amp; WT have more prestige than those who do not.</td>
</tr>
<tr>
<td>IMG4</td>
<td>I use AR &amp; WT solutions, because I want to be a forerunner in technology exploitation.</td>
</tr>
<tr>
<td>IOP1</td>
<td>Interoperability is important for AR &amp; WT.</td>
</tr>
<tr>
<td>IOP2</td>
<td>I am worried about vendor lock in with AR &amp; WT.</td>
</tr>
<tr>
<td>IOP3</td>
<td>Integration costs of AR &amp; WT with other software systems in use are high.</td>
</tr>
<tr>
<td>IS6</td>
<td>I would find it useful if my friends knew where I am and what I am doing.</td>
</tr>
<tr>
<td>LRN1</td>
<td>Learning curve for AR &amp; WT is too high compared with the value they would offer.</td>
</tr>
<tr>
<td>PE10</td>
<td>With AR &amp; WT, I immediately know when a task is finished.</td>
</tr>
<tr>
<td>PE4</td>
<td>Using AR &amp; WT increases my productivity.</td>
</tr>
<tr>
<td>PE8</td>
<td>AR &amp; WT increase precision of tasks.</td>
</tr>
<tr>
<td>SI1</td>
<td>People who are important to me think that I should use AR &amp; WT.</td>
</tr>
<tr>
<td>BI2</td>
<td>I will always try to use AR &amp; WT in my daily life.</td>
</tr>
<tr>
<td>UF1</td>
<td>Please choose your usage frequency of AR/WT</td>
</tr>
</tbody>
</table>

The model explains how accepting of AR/WT technologies are questionnaire respondents, and predicts their intention to use them, and likely future usage behaviour. In the validation trials at ALTEC, **199 participants** filled in the questionnaire.
The figure below depicts this analysis visually: The ‘hinges’ in the diagram cover the answers of 50% of the participants, with the line in the middle indicating the median. The ‘whiskers’ span the 95% confidence interval. Outliers are depicted with a black dot. The connected red dots indicate the means (Fig. 8).

**Figure 8.** TAMARA results for the Space trial

Considering the 7-point Likert scale ranging used, it is possible to assess that a mean of 5.13 with a standard deviation of 1.28 indicate a good acceptance of the technology.

### 3.1.2. System Usability Scale (SUS)

System Usability Scale (SUS) is a tool for measuring both usability and learnability. The SUS scores calculated from individual questionnaires represent the system usability. SUS yields a single number representing a composite measure of the overall usability of the system being studied. Scores for individual items are not meaningful on their own. SUS scores have a range of 0 to 100 (Brooke, 1996;
According to validation studies, the SUS score starting from 68-70 represents the level of acceptable system usability. The Acceptability ranges are: 0-50 not acceptable; 50-70 marginal; 70-acceptable (Bangor et al., 2009; Brooke, 2013).

<table>
<thead>
<tr>
<th>Applications</th>
<th>SUS score</th>
<th>Applications</th>
<th>SUS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player (P) users only</td>
<td>69</td>
<td>P + P and R users</td>
<td>68.39</td>
</tr>
<tr>
<td>Recorder (R) users only</td>
<td>53</td>
<td>R + P and R users</td>
<td>56.68</td>
</tr>
<tr>
<td>Overall</td>
<td>59</td>
<td>Overall</td>
<td>67.64</td>
</tr>
</tbody>
</table>

The SUS score for Player (69) has reached the range of acceptable system usability. The SUS score for Recorder (57) represents the marginal range. The number of test participants using the Player was significantly higher (147) than the number of Recorder users (22), thus the comparison between Player and Recorder scores is only approximate. The acceptability range result for Player can be seen as valid because of the high number of users (N=169). The acceptability of the software at ALTEC is rated higher than in the other pilots (LT: 67; EBIT: 61).

3.1.3. Spatial Interface Evaluation (SPINE)

Spatial Interface Evaluation (SPINE) was developed by partners from OBU specifically to evaluate usability of the augmented reality user interface. This questionnaire was developed and designed specifically to evaluate those usability aspects that the standard and industry-established SUS method is missing.

While reading the graph it is worth considering that questions MP2, MP3, MP4 and IM1 were negatively formulated and should be read with reverse polarity in mind (see Figure 10 for the questions). Considering the 7-point Likert scale ranging used, it is possible to assess that a mean of means of 5.04 at a standard deviation of 1.3 indicates a good usability.
The figure above depicts this analysis visually. The 'hinges' in the diagram cover the answers of 50% of the participants, with the bold line in the middle indicating the median. The 'whiskers' span the 95% confidence interval, indicating minimum and maximum values excluding outliers. Outliers are depicted with a black dot. The connected red dots indicate the means (Fig. 9).

The hinges (50% of the participants) of the items NG3 (Orientation - I had difficulties to identify in which direction I need to face), MP2 (Accuracy - The user interface interrupted my real-world task performance) and MP3 (Precision - The interface blocked my interaction with the real world) had the widest upper-quartile and lower-quartile spread.

Overall, participants agree that the system helped them to accomplish the training aims and the results suggest that the system has a good usability. They clearly indicated that they understood functionality they have used (SL2) and it was easy to control the system (see SC group). The only negative aspect that can be highlighted is that participants somewhat agreed that they had difficulties to identify the right direction they need to face. But at the same time, in NG1 (Median value "Somewhat agree") they pointed out that "It was easy to find the target with the navigational pointers provided". The results are similar to the EBIT case.
**Wearable Experience for Knowledge Intensive Training**

---

**Figure 9. SPINE questions: 6 groups of questions in total.**

The SPINE method consisted of 29 questions group into 6 categories (System Control, Navigation, Manipulation, Selection, Input Modalities and Output Modalities). Each of the categories is looking into a specific aspect of its domain, i.e. Navigation is exploring usability issues related to location, obstacles, orientation and wayfinding. This data together with SUS and TAMARA will help to conduct a further and more detailed regression analysis.
3.1.4. Knowledge Assessment Results

The aim of the Knowledge Assessment test was to evaluate participants' performance after the training. Experts at ALTEC designed the test, and almost all the knowledge test questions are testing knowledge acquired during corresponding procedural steps. In total, there were 15 procedure steps and 15 knowledge test questions.

![Knowledge Assessment (Space)](image)

**Figure 11.** Knowledge Assessment results (AR condition vs Control group); see Table 1 for questions

Mean AR group: 66%, Mean Control group: 63%.

The figure above depicts the analysis visually. The black dots and lines represent AR group, and red - control group, the values (in %) indicate the correct answers (see Table 1 for questions).

**AR group** completed 66% of the questions correctly. One question was answered 100% correctly: AL9 (Which colour is the rover battery charge connector?). **Percentage of correct answers for the questions AL12 (Which warning was triggered during the procedure?) and AL15 (How will you reactivate the rover at recharging procedure completed?) was below 50% (11% and 14% respectively).**
Control group completed 63% of the questions correctly. Four of the questions were answered 100% correctly: AL2 (On the rover control panel picture provided, please identify the correct location of the red key), AL4 (On the rover picture provided, please locate the red safety buttons), AL13 (What did you do after alarm?) and AL14 (How would you move to the following procedure step?). Percentage of correct answers for the questions AL12 (Which warning was triggered during the procedure?) and AL15 (How will you reactivate the rover at recharging procedure completed?) was below 50% (25% and 15% respectively).

![Z-score for testing equality of proportions (Space)](image)

**Figure 12. Z-score for testing equality of proportions between AR group and Control group**

Z-test was calculated for the hypothesis of equality of proportions of unconditional means between AR group and control group. The results show there is no statistically significant difference between unconditional means. Regression analysis is required for the more detailed results. Knowledge assessment of the AR group and Control group showed that Control group performed better on Question AL1, AR group performed better on Question AL9, but this did not influence the overall result (ALP with Z-score 0.3 and p-value 0.76).

3.1.5. Expert Evaluation

The aim of this expert model evaluation questionnaire based on Jucks, Schulte-Löbbert, & Bromme (2007) was to peer-evaluate the expert model (recording of the expert), by judging its fitness for the purpose via other experts (recording of the expert). The participants responded by selecting from 1 (strongly disagree) to 7 (strongly agree). This expert model was facilitated by the affordances of the
wekit.one Recorder application and therefore, co-relates to the extent to which the application enables effective recording of the expert. 22 experts participated in the study, however, after removing those participants who did not use the player and those who did not come directly from the domain, 13 expert participants remained. There were 2 female and 11 male expert participants, with majority of them falling in the age range of 25-34. Among these participants, there were 1 supervisor, 1 trainer, 1 trainer and supervisor, 9 engineers and 1 system administrator. While all the expert participants did not directly have teaching role in the domain, they did however have good experience in field. 7 expert participants had more than 10 years of experience, 4 between 5-10 years and only 2 less than 5 years.

Figure 13. Comparing average response in expert participants for each item

The expert participants mostly responded positively to the model with the mean and median of all items above 4. Most expert participants strongly agreed on EMEQ1 and rated the item between 5-6 with only 1 participant rating it 4. EMEQ 2 was rated between 4-7 by most participants, which verified that most concepts were defined just in time by the expert model. Participants also strongly agreed to score EMEQ 3 between 5-6 with only 1 expert participant rating it 4. In item EMEQ 4, most students rated the item between 4-6. These results show that the expert model explained the procedure in comprehensible terms and details. Results of item EMEQ 5 show that the contained information in the expert model is complete. EMEQ 6 was rated between 3-7, showing that the expert participants’ agreement varied in terms of if the information was provided in the right time. EMEQ 7 and EMEQ 8 were rated between 4-7 by the participants with only 1 participant rating EMEQ 8 below 4. It validates that the expert model contains all the important information, which is presented in an unobtrusive manner. Item EMEQ 9 was rated between 3-6. Participants were fairly able to locate the objects required for the procedure most of the time. In item EMEQ 10, there was a strong agreement between the expert participants with most participants rating it between 5-6 which verifies that the participants were always able to identify the place where the next step of the procedure was to be done. EMEQ 11 was rated between 2-6 with only 75% of the sample voting it between 4-6. While the score can be interpreted to state that the important information was fairly frequently updated by the
model, this aspect can be affected by various factors such as the notification being updated was overlooked, or that there was no information that needed to be updated. However, average results show that the expert participants found the expert model created by the wekit recorder and replayed by the wekit player to meet the requirements of an expert model. This model was peer reviewed by the expert participants without having to perform post analysis on the recorded data.

<table>
<thead>
<tr>
<th>Table 4. Means and Std. Deviation of the experts’ response</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>EMEQ1</td>
</tr>
<tr>
<td>EMEQ2</td>
</tr>
<tr>
<td>EMEQ3</td>
</tr>
<tr>
<td>EMEQ4</td>
</tr>
<tr>
<td>EMEQ5</td>
</tr>
<tr>
<td>EMEQ6</td>
</tr>
<tr>
<td>EMEQ7</td>
</tr>
<tr>
<td>EMEQ8</td>
</tr>
<tr>
<td>EMEQ9</td>
</tr>
<tr>
<td>EMEQ10</td>
</tr>
<tr>
<td>EMEQ11</td>
</tr>
<tr>
<td>Valid N (list-wise)</td>
</tr>
</tbody>
</table>

### 3.2. Interviews results

Additionally to the questionnaires, seven participants were interviewed during the space case trial. These interviewees only used the Player. All the interviewees were working in ALTEC and were considered as experts. Five interviewees were male and two female. Interviewees were chosen randomly based on their own motivation and their possibility to spend time for the interview.

The following themes and questions were used in the interviews:
Please use 2-3 words (adjectives) to describe your experience with the AR system (Player & HoloLens) during the execution of the task. Choose the words based on your immediate response.

- Please describe freely your experience on using the AR system during the execution of the task.
- How well did you manage to complete the task with the player?
- Please tell the good and bad sides (pros and cons) of AR glasses (HoloLens). 2-3 positive and negative aspects.

The summary of the interview results is presented below.

**Results - overall experience**

The interviewees were quite satisfied with the whole AR system (HoloLens + Player):

<table>
<thead>
<tr>
<th>Positive expressions on the experience with the system</th>
<th>Negative or neutral expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interesting (3 times), nice, intuitive, easy to use, useful, innovative, immersive, enjoyable, simple, instructive, helpful, clear, smart, I felt more confident with the task</td>
<td>upgradable a bit, strange, difficult</td>
</tr>
</tbody>
</table>

Comments related to HoloLens device only:

<table>
<thead>
<tr>
<th>Positive expressions</th>
<th>Negative or neutral expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>It’s light (expected to be heavy); field of view was good; comfortable; easy to adjust on the head; very clear images; quite comfortable to wear; AR is perfect for this kind of task - better than VR; well integrated with life support box; immediate; wearable; it can ease a lot different procedures; I used with my eyeglasses - no problem with that; wearable - you can adjust it and it stays in the head</td>
<td>Too heavy (2 times); device could be lighter; a little bit tricky to wear with eyeglasses; I don’t know how it works in collaboration with many people; I felt a little disoriented during the task</td>
</tr>
</tbody>
</table>

**Results - usability and UX issues**

Many interviewees reported difficulties in understanding what is the right/left or rear/front of the rover. A better reference system (of the rover) was hoped for. Positioning the menu/tasklist was difficult and test users felt its position sometimes difficult and uncomfortable. They didn’t know or remember the LOCK/RELEASE MENU function.

The line guidance was sometimes difficult to follow/understand. Test users got lost sometimes. This seemed to happen more often in the beginning of the task, when the users were not familiar with the system guidance yet.

All the interviewees used only or mostly voice commands, and they were happy with them. One user reported problems remembering the voice commands. According to one user, the tapping (clicking) gesture was difficult and the person was not able to do it.
Some users reported that instructions were a little bit misleading or unclear sometimes: The red button was not that “BIG” as described (emergency case) or the control panel was expected to be bigger. Generally, the quality of instructions was considered as good.

Many users reported that the beginning of the task was difficult, but after learning and understanding how the system works, it became easier. Since there is no actual feedback on user performance, some users felt uncertain if they had performed the step correctly.

### 3.3. Biosensor Data

The WEKIT biosensor harness (BH) was used to measure trial participants’ physiology, specifically, blood volume pulse (BVP), galvanic skin response (GSR), bodily and environmental temperature and humidity, and two sets of inertial measurement sensors comprising accelerometer, gyroscope and magnetometer readings. A total of 42 trials involving the BH in conjunction with the Hololens were conducted across the Space and Aerospace cases, with the BH generating time-stamped datasets, which were then manually paired with questionnaire responses.

Both participants and facilitators found the BH easy to work with in comparison with earlier attempts using the sensor-integrated jackets or vests. ALTEC facilitators had an issue with the battery fatigue that prevented data acquisition for some of the trials, however this was resolved through adapting the charging protocol.

On post-analysis it was found that some of the datasets contained noticeable gaps where the wireless UDP connection temporarily failed. Such gaps often recovered quickly but must be accounted for when computing time-series statistics so as not to influence the higher-order variables.

Whilst studies exist establishing links between physiological readings and affective state (see D4.4) they are usually performed in controlled lab conditions. Obtaining such consistent data in the WEKIT trials, particularly for BVP and GSR, proved difficult since participants were required to perform a number of full-body manual tasks. However, manual inspection of the datasets revealed regions of high-quality data for every one of the BH trials, and such regions of interest could be selected for further work using signal-processing methods not employed here.

Various methods for measuring the pulse of moving subjects were considered in D4.4 and electrocardiogram (ECG) was found to deliver the most consistent data, yet it suffered the drawback of requiring adhesive single-use pads to be placed intrusively on the body. Potential future studies taking physiological readings in industrial trials may still find ECG to be the preferred method over BVP.

### 4. Discussion

#### 4.1. Quantitative Results: Questionnaires

**4.1.1. Technology Acceptance Model for AR/WT (TAMARA)**

The Technology Acceptance Model for AR/WT questionnaire pointed out that overall the participants are acceptant of the novel technology they tested (mean of means: 5.13 with a standard deviation of 1.28), see section 3.1.1. The majority of participants clearly (with a median of 6) expressed a positive
attitude towards the technology (ATU4) and show signs of hedonic motivation to work with the technology (HM2b), expect low effort in working with the technology (EE2), and are positive about the availability of facilitating conditions and resources required (FC1). They are also positive about statements such as: importance of interoperability of AR & WT (IOP1), the support given by the system allowed [them] to perform the procedure to a higher standard than without (ITF1), with AR and WT I immediately know when a task is finished (PE10), using AR & WT increases my productivity (PE4), this particular procedure was well supported by the system (ATF1), AR & WT increase precision of tasks (PE8), in the future, I would feel comfortable sharing the personal data captured with my organisation (IS7), the information provided by the system was appropriate for my level of expertise (ITF2) and The right amount of information was presented in each step of the activity (ATF2).

They in general express a tendency to agree that the technology expresses a high degree of computer self-efficacy (CSE4). The participants consider integration into the wider system landscape as a desideratum (IOP1). Their acceptance with respect to habits (HT2), existing image gains within the organization (IMG1), is, however, neutral. Furthermore, they do not fear vendor lock in (IOP2). They are similarly unsure about the facts that people who are important to them think that they should use AR&WT (SI1) and daily usage of AR&WT in their life (BI2).

4.1.2. System Usability Scale

The System Usability Scale (SUS) questionnaire showed that the SUS score of 68 was very close to 70 (scale 0-100), which is considered as the minimum acceptable score for system usability. Thus, the WEKIT applications were considered nearly acceptable in terms of system usability, the participants' feedback indicated that they would be disposed to use the WEKIT applications for performing their activities since the system is not too complex and the features are well integrated and consistent.

Furthermore, participants somewhat agreed that they would need the support of the technical person to be able to use the system. On it is own it very hard to explain this fact, that's why we have used SPINE method to determine the problem. At the same time, participants agree that they felt very confident using the system. Additionally, they disagree with the fact that there was too much inconsistency in the system.

4.1.3. Spatial Interface Evaluation (SPINE)

The Spatial Interface Evaluation showed a good usability of the technology. Overall, participants agree that system helped them to accomplish the training aims and the results suggest that the system has a good usability.

The feedback clearly indicated that participants understood functionality they have used. All items in the System Control group have a median value of 6: it was easy for them to control the application (SC1 - Ease of Use), user interface elements clearly represented their functions (SC2 - Metaphors), participants were able to understand where they are in the system (SC3 - Status) and the user interface blended together with the real world (SC4 - Realism). In addition to that Selection group has also a median value of 6: the size of the UI elements was sufficient to select (SL1 - Size), participants understood the functionality they used (SL2 - Instruction), the system provided timely feedback about the processes happening (SL3 - Feedback) and the interface provided clear confirmation of the actions participant performed (SL4 - Confirmation).

The only slightly negative aspect that can be highlighted is that participants somewhat agreed that they had difficulties to identify the right direction they need to face (NG3 - Orientation). However, at the same time, in NG1 - Location (Median value of 5) they pointed that “It was easy to find the
target with the navigational pointers provided” and in NG4 – Way finding (Median value of 6) that “It was clear how I need to position myself to have the best experience”. Further regression analysis will point out towards the connection between different groups, but at this moment it is clear that SPINE methods provide us with more insight regarding the user interaction with the interface and system usability aspects that the SUS method. Specific aspects are now identified (Orientation) and can be considered in the future.

4.1.4. Knowledge Assessment Results

The knowledge assessment test showed no statistically significant difference between AR group (completed 66% of the questions correctly) and control group (completed 63% of the questions correctly). Overall, the pattern of the correct answers is similar. It is worth noting that there were no procedure step assigned to knowledge test items AL12 (Which warning was triggered during the procedure?), AL13 (What did you do after alarm?), AL14 (How would you move to the following procedure step?), and both AR group and Control group scored low and high on these questions (see Figure 14). Consequently, there were no associated media content to these questions, and knowledge is accumulated either as a synthesis of critical thinking or part of the observation.

To conclude, the results of the knowledge test do not indicate that AR technology is performing better within the selected sample group and it also suggest that AR technology is not worse than classical training methods. The further analysis will reveal what influences the performance within two groups - age, gender, education, prior experience or maybe they way participants answered to TAMARA/SPINE questions.

**Figure 14.** Knowledge Assessment results (AR condition vs Control group); see Table 1 for questions

Mean AR group: 66%, Mean Control group: 63%.

Additional considerations can be done from the technical point of view. The small difference assessed in the score results between the test subjects who used the AR software and the ones who were part of the Control Group is probably also due to the fact that astronauts procedures are normally very
reliable and clear, since they are developed according to specific standards. In fact the SSP 50253 Operations Data File Standards (latest issue available December 2017, Rev Y), an International Space Station Program document, is written and maintained with the following purpose:

“defines the formats and conventions used in producing Operations Data File (ODF) procedures. This document contains standards that apply to all ODF procedures and standards that are unique to a single Partner’s ODF procedures. All ISS Partners involved in ODF procedure development will abide by the standards documented in this document.”

“The Operations Data File (ODF) is the collection of ODF procedures and reference information. An ODF procedure is a set of instructions used by specially trained personnel such as ground controllers, the on-board crew, and the on-orbit procedures executor software to fulfil certain tasks. These are the tasks needed to operate and maintain station systems, payloads, and attached vehicles under both nominal and off-nominal conditions.

Standardizing the format of the procedures used on-board the Space Station by an international crew is necessary for consistency because the crew moves from one element to another, and from one activity to another. The standards for developing ODF procedures to which all program participants shall conform are described.”

Scope of this standardization is to develop procedures as complete and clear as possible, to make them understandable by all crew members, avoiding mistakes as much as possible. The “paper procedure” used by the Control Group was developed according to these standards so the task was well understood and learnt by all participants.

The usage of the WEKIT prototype, however, was able to slightly improve the means for the already good knowledge retention result, demonstrating that the use of AR, beside been beneficial on the short term during the task execution, is also affecting the participants learning curve.

4.1.5. Expert Evaluation

The results from the expert evaluation show above average quality of the expert model with fairly consistent agreement among the expert participants in their response. One expert who was well familiarized with the technology, the teaching procedure, and the recording methodology created the expert model. He was given ample time and assistance for recording the expert model. The recording of the expert model was kept as authentic as possible, using the functionality of the recorder and limiting post processing to simple actions such as recompressing images, which would not alter the recording by any means. The recorded model was peer evaluated by other experts by using the player. The evaluation was done based on the items proposed by Jucks, Schulte-Löbbert, & Bromme (2007) and adopted for the WEKIT scenario, which checks if the expert model meets certain criteria that makes it comprehensible for the students. The expert model seems to have met the scores for the items in this questionnaire, with a notably low score in the EMEQ 11, which evaluated the suitability of the updates for the relevant information. This information could have been sensor-based information or just notifications on change of state which are important for the students.

4.2. Interviews

Based on the interviews (N=7) on the use of Player and HoloLens, it can be stated that

- The overall experience with the AR system was considered as satisfactory
Learning how to operate the AR system seemed to improve quite effectively during the task execution, feedback however should be given on users' performance

- The line guidance for finding the objects and places should be instructed and visualised more carefully
- The critical points of the rover (left/right/rear/front) should be instructed more clearly and access to e.g. a picture of the rover should be provided when needed

4.3. Biosensor

The WEKIT Biosensing Harness was successfully developed and utilised in the Space case industrial trials across a diverse set of participants over a period of several months. From this we have obtained a complex, but valuable dataset using a bespoke wearable system, and in our initial analysis have developed methods for overcoming data corruption that occurs as a result of measuring the physiology of moving subjects. Further analysis may reveal links between biometric signals and learning outcomes or even perceptions of AR/WT. Methods for sensor-fusion pattern detection are suggested in D4.4 but have not been implemented here.

5. Conclusions

The objective of this second and final trial for the space pilot case was to test the hardware and software prototype on a real procedure used for the astronaut training activities, as defined in D6.9 - Training Scenario and Evaluation Plan for Space.

The procedure identified and described in the deliverable (Recharge of batteries of a Mars Rover prior to perform an exploration mission) was carefully analysed and detailed, to identify the best AR solutions that could improve the complete understanding of each step by the test participants, who were only partially coming from the space business (so familiar with the constraints and the difficulties to be faced). In fact, most of the participants to this second part of the trials have been selected among "common" people, without a specific background in space operations.

The number of participants was high. The impression collected during the tests, however, was that the reactions were more varied than expected: people with similar age and technical background were showing quite different performance. The feeling was that training could reduce this variability, but proficiency and performance will always be more varied than expected.

The effort dedicated to organize and perform the trials was very demanding and required the involvement of three ALTEC instructors from July 2018 up to February 2019, to support and follow the testing of the WEKIT prototype executed in ALTEC by 199 participants in total.

All the participants provided positive feedback about the usage of the tool that was appreciated because it demonstrated how the task execution and the learning process was improved in terms of effectiveness, time reduction, and user perception by using the WEKIT methodology.

The ALTEC trainers had the possibility to test the prototype both from the player and recorder perspective, having in mind its potential utilization in the crew-training environment. The astronaut training process needs to evolve taking into account the need to adapt to the different conditions of exploration missions. The duration of the training period and the communication delays, that will be one of the major problems to be faced to prepare the astronauts for mission on Moon or Mars, highlight the necessity to make astronauts more independent and autonomous in executing their
tasks and also in reacting to contingency and even emergency events. The WEKIT hardware and software prototype is expected to fill these gaps by reducing the hours of training dedicated to maintenance activities and providing support to the crew when communication with Ground is not available.

Most of the experts involved were ALTEC and Thales Alenia Space personnel, who already participated to the first pilot back in May 2017. They had the possibility to compare the prototype evolution during the project. The use of the Hololens to perform this kind of activities made the experience more appealing even for the ones who had already tried Virtual and Augmented Reality Technologies. They understood the potential of applying such a technology to this field and they were very collaborative in giving advice and feedback on how the prototype could be further enhanced.

The students were very enthusiastic to test such a tool on a real procedure. The combination of the Hololens and the Martian Rover, located in the ALTEC Mars Terrain Demonstrator, where the test was performed, made the experience unique for them.

Some of the problems highlighted during the test execution were related to the usage of the Hololens: For example, the limited-field-of-view experienced by some testers and the battery problems described in para 1.5.3 are consequences of using version 1 of the HoloLens. Pre-release information on version 2 suggest that the above problems will be largely solved (i.e. HoloLens 2 with a wider field of view and more precise hand and environmental tracking).

The use of the Lime Survey tool to complete online the questionnaires was a very good idea to quickly collect feedback from all participants and also to perform the data analysis on the large number of participants. Overall, participants suggest that they felt confident using the system and are open minded about using this technology in the future. The general acceptance and usability of the WEKIT tool is good and that suggest us a promising perspective of its future.

Finally the utilization of the “WEKIT biosensor harness” allowed to collect data used to measure trial participants’ physiology, in order to evaluate potential “stress” indicators that could highlight procedure steps not completely clear or difficult to perform for the participant, despite the AR information included to clarify as much as possible the context.

In conclusion, ALTEC are particularly interested in continuing the exploitation of the WEKIT tool and their intention is to propose its utilization to enhance the training process for astronaut and to support future Exploration missions on Mars or on the Moon, proposing to the major Space Agencies to perform further testing with crew members and scientists in the so called “analogue” environments (like PANGAEA and CONCORDIA).
References


