D5.2
VALIDATION
WITH STUDENTS

Approval Status

	NAME AND SURNAME	ROLE IN THE PROJECT	PARTNER
AUTHOR(S)	F. Morosi	Researcher	POLIMI
	I. Carli	Researcher	POLIMI
	G. Caruso	WP2 Leader	POLIMI
	J. O'Hare	Researcher	UBATH
	F. Ben Guefrache	Researcher	GINP
REVIEWED BY	J. O'Hare	Researcher	UBATH
	G. Caruso	WP2 Leader	POLIMI
	G. Cascini	Project Coordinator	POLIMI
APPROVED BY	G. Cascini	Project Coordinator	POLIMI

History of Changes

VERSION	DATE	DESCRIPTION OF CHANGES	BY
01	07.06.2018	Initial draft of the document	G. Caruso
02	18.06.2018	Hypotheses and metrics	J. O'Hare
03	19.06.2018	Treatment of data	J. O'Hare
04	20.06.2018	Colour accuracy	G. Caruso
05	27.06.2018	Description of the different interfaces	I. Carli
06	28.06.2018	Description of the experimental protocol	I. Carli
07	29.06.2018	Elaboration and discussion on the log File	F. Morosi
08	30.06.2018	Participants' perception of usability	J. O'Hare
09	02.07.2018	Introduction, graphs for Accuracy of placement, rotation and scaling of assets for Time	J. O'Hare
10	04.07.2018	Review of the document	J. O'Hare

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

11	05.07.2018	Elaboration and harmonization of all images and graphs	I. Carli
12	06.07.2018	Comments on efficiency	F. Morosi
13	10.07.2018	Comments on accuracy	I. Carli
14	11.07.2018	Expected hypothesis	F. Ben Guefrache
15	12.07.2018	Conclusions	F. Morosi
16	13.07.2018	Review of the document	G. Caruso
17	14.07.2018	Final revision	G. Cascini

Document Details

DISSEMINATION LEVEL	Public
DUE DATE	30.06.2018
ISSUE DATE	15.07.2018
CONTRACT NUMBER	H2020-ICT/2015-688417
ELECTRONIC FILE LOCATION	http://www.spark-project.net/wp-deliverables
FILE NAME	D5.2_WP5_Validation_with_students

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

I. TABLE OF CONTENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.

H2020_ICT_2015_SPARK_688417

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

II. LIST OF FIGURES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

1 EXECUTIVE SUMMARY

This deliverable reports the activities carried out in task 5.2 "Validation with students." The objective of these activities is to present the SPARK project to an audience of students by involving them in using the SPARK platform to evaluate the effectiveness and the usability of the different User Interfaces (UI) implemented for the SPARK platform. The document starts with the description of the aims related to the task 5.2 and introduces the metrics used for the evaluation tests. Then, the different UIs for the SPARK platform are presented with the hypotheses formulated in relation to the aspects we intended to evaluate.

The activities of the tests, reported in the document, involved students from the three academic partners of the SPARK consortium (POLIMI, GINP, UBATH). Each institution organized the tests in their premises with the aim, also, to evaluate different technical solutions for the SPARK platform. Students have been invited to replicate different graphical layouts of a product, which was used within the activities of WP4.

The test included objective and subjective evaluation. A logging module has been expressly included within the software of the SPARK platform to monitor the students' activities. Whilst, specific questionnaires, which are reported in the appendix of the document, have been proposed to the students at the end of the test execution.

The analysis of the data collected during the tests shown that all students managed to use the SPARK platform without particular issues and the performance is similar, and sometimes better if compared with more common interaction systems. The positive judgments in terms of usability confirm the rightness of the choices made for the implementation of the UI of the SPARK platform. The results and all the observations done in the accomplishment of task 5.2 represent valuable insights for the refinement of the last version of the platform, which is due by the end of M31.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

2 INTRODUCTION

This deliverable describes the activities completed within T5.2 – Validation with Students. The objectives of this task were to:

- Gather feedback on the usability of the SPARK platform in order to inform the final development activities (T3.2) as well as the long-term evolution of the platform.
- To introduce the SPARK platform to potential future users and customers.
- To trigger interest amongst students in applied research in the domain of engineering design.

It was originally foreseen that the validation with students would involve student groups using the SPARK platform within co-creative sessions to work on realistic design briefs set by the industrial partners. Based on the consortium's experience of running these types of sessions within the experimental activities of WP4, it was felt that there was little value in running this type of session because:

- The data collected would be less 'valid' than the data already obtained in WP4 that involved experienced design practitioners working on real projects.
- The feedback from design practitioners during WP4 had already validated the concept of the SPARK platform but had highlighted the need to make significant improvements to the usability of the system.
- The complexity and dynamics of working on an open-ended design task within a co-creative session would make it very difficult to gather high-quality feedback on the usability of the SPARK platform or compare the usability of different user interface configurations.

It was therefore decided to change the experimental task to enable a greater focus on testing the usability of the SPARK platform. Specifically, it was decided that the task should be completed individually and should concern the replication of the graphic layout of a printed packaging product. Working individually eliminated any risk that the user feedback would be influenced more by the success of the group collaboration performance than by the usability of the SPARK platform. Likewise, the task of replicating an existing design eliminated any risk that the user feedback would be influenced by the perceived creativity of the session output rather than by the usability of the SPARK platform.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

During the development of the SPARK platform, a number of different alternative user interface configurations were proposed. However, WP4 tests provided limited feedback on the relative performance and efficiency of the alternative user interfaces. Hence, a key objective of the student tests was to compare the performance of these alternatives such that one could be selected for further development. Further details of the alternative user interfaces are provided in Section 3.1.

A variety of different metrics was applied in order to assess the performance and usability of the alternative user interfaces. These included measures to assess:

- The **accuracy of the placement, rotation and scaling** of graphical assets placed on the packaging.
- The **accuracy of colour selection** for the background of the packaging.
- The **efficiency** of the system in supporting the rapid completion of the task.
- The users' perception of the **usability** of the system.

Section 3.2 presents further details of the metrics and associated hypotheses.

Tests were completed at the premises of each of the three academic partners. This allowed for a greater number of student participants to be recruited but required careful planning and coordination to ensure the comparability of the results. Further details of the experimental protocol and participants are provided in Section 4, whilst the results are presented in Section 5. General conclusions, including recommendations for future development of the SPARK platform, are presented in Section 6.

3 EXPERIMENTAL CONDITIONS AND HYPOTHESES

3.1 DESCRIPTION OF THE DIFFERENT INTERFACES

The User Interface (UI) developed for the SPARK platform has been set in five different conditions (A, B1, B2, B3, B4), which are the subjects for the usability testing activities performed with students. Each user, participating in the experiment, worked with condition A and afterwards with at least one among the conditions B. Further details on the experimental protocol are provided in Section 4.

The UI functionalities that include browsing of the asset library, selection of the 3D model part and some functionalities related to the asset editing (change of layer, asset delete and replacement) remain

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

unchanged among all the conditions because they are activated by pressing UI buttons. In the following, all the conditions are described in detail.

Condition A, as illustrated in Figure 1, uses traditional mouse and keyboard for the interaction while the monitor screen is the visualisation system that shows the Graphical User Interface (GUI), as visualisation input, and allows the user to evaluate the 3D model, as visualisation output.

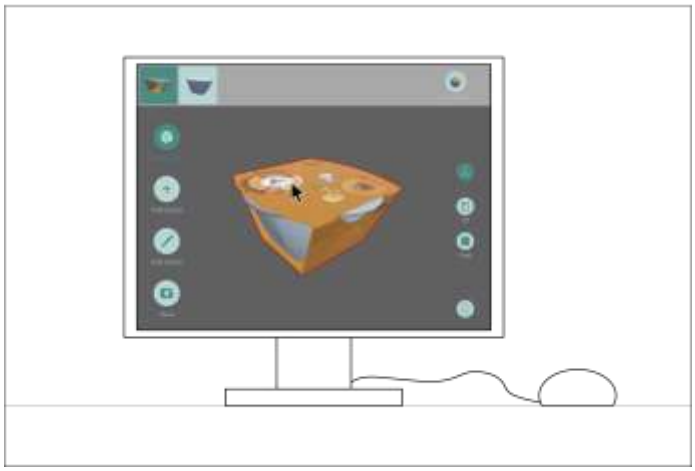

Figure 1: Condition A and B4

In **conditions B1, B2, B3** (Figure 2) a multi-touch tablet is used for the interaction (tablet model: Samsung Galaxy Tab S2 (2016) (9.7")) and as visualisation input. The multi-touch gestures implemented to modify the layout of the prototype are *"pinch"* to scale an asset, *"rotate"* two fingers clockwise and counterclockwise to modify the orientation of the asset, *"drag"* a single finger to move an asset. The visualisation output in all the three conditions is the mixed prototype. It consists of a white painted mock-up where the projection on its surface allows the user to see in real time the modifications that she/he is performing on the product.

In **B1**, the GUI includes a 2D outline representing the UV map of the corresponding part of the 3D model, as illustrated in Figure 2a. In this configuration, the user sees all the assets layered on the UV map and has a general overview of the final configuration of the layout.

In **B2**, the GUI includes an empty area where the user can modify the selected asset through the implemented multi-touch gestures (Figure 2b). The tablet UI is only showing the menu buttons and the asset library while the real-time modification of the assets can be only viewed on the mixed prototype.

In **B3**, the GUI includes the view of the 3D model, as in condition A, with the difference that the user makes changes on the assets with multi-touch gestures (Figure 2c).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

Condition B4 is equal to **condition A** and uses the traditional mouse and keyboard for the interaction and the monitor screen for the visualisation (input and output), as shown in Figure 1. The aim of this condition is to keep control of possible learning effects due to the use of **condition A** in the first round of the test.

Figure 2: a. UV Map (B1), b. Touch Area (B2), c. 3D View (B3)

3.2 HYPOTHESES AND METRICS

In Section 2, it was noted that the main objectives of the tests with students were to assess the performance of the SPARK platform in terms of its accuracy, efficiency and usability and to compare the performance of the alternative user interface configurations described in Section 3.1. Table 1 provides a summary of the metrics and their definition.

TABLE 1. SUMMARY OF THE METRICS APPLIED AND THEIR DEFINITION.

Metric	**Definition**
Placement accuracy	Displacement of the placed asset from the reference location[1].
Rotation accuracy	Angular displacement of the placed asset from the reference angle[1].
Scaling accuracy	Difference between the scaling factor of the placed asset and the reference scaling factor[1].
Colour accuracy	Colour difference between the colour selected by the user and the reference colour printed on the template. Measured in terms of lightness, Chroma and hue, distance calculated using the *CIEDE2000* colour-difference formula.

[1] The reference location for an asset was calculated as the mean x and y coordinate for that asset in the specified layout (across all conditions). The reference location was calculated separately for each asset and for each test location. This ensured comparability of results in spite of any differences in projector alignment between the test locations. The reference angle and reference-scaling factor were calculated in a similar manner.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

Efficiency	Time dedicated to complete the task.
Usability	Usability score calculated from the results of the System Usability Scale and Creativity Support Index surveys.

The measurement of the errors in the asset placement, rotation and scaling is relevant for the evaluation of the accuracy of the proposed interfaces based on the interaction with the touchscreen. In fact, these modifications are the only ones applied with different interaction modalities with respect to the mouse (Section 3.1), which is considered a relevant benchmark since it is widely used and it guarantees a good precision. We expect to have a significant difference between the two phases of the tests (condition A and B) due to the learning effect (visible in condition B4) and the lower precision provided by the touch (visible in condition B1, B2 and B3). In addition, due to the differences in terms of system and capability of the tracking system between the UBATH and POLIMI, we expect to meet different performance.

The colour similarity between the background of the projected prototype and the real one is used to check the calibration procedure adopted to improve the rendering quality of the projectors. By using the hypothesis described in Section 4.5.2, we analysed the discordance between the colour perceived by the user and the real one. Due to the possible human misperception of projected images, our expectation is to have a relevant discordance of the colour selected by the user.

The estimation of the efficiency of the interface is done with the measurement of the total time used to complete the task, which can represent an effective way to evaluate this aspect. We defined the time interval as a not mandatory time limit within which all the modifications should be performed. Due to the major complexity of interaction, condition B1, B2 and B3 are expected to be more time consuming than the A, as well as B2 among those using the touchscreen.

The overall usability evaluation is performed thanks to the completion of the SUS and CSI surveys, which allow better understanding subjective aspects. According to the tests performed in WP4, our expectation is to obtain a higher score in condition B1 and B3 than in condition B2, but a similar mark with the mouse condition.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

4 EXPERIMENTAL PROTOCOL

4.1 DESCRIPTION OF THE TASKS

In Section 2.1, it was explained that the focus of these tests was on comparing the usability of the alternative SPARK user interface options. It was, therefore, necessary to develop a task that would test the usability of the system without the risk of the influence of possible external variables – such as the complexity of working in a co-creative design setting, or the differences in creative design abilities of the participants. It was therefore decided that the task would be completed individually. The task involves replicating the external graphic layout of a printed packaging product, as one of those used in the WP4 testing activities.

In particular, the product is a soup container, composed of two main components: the external cardboard sleeve and the plastic bowl. Four alternative layouts of the cardboard sleeve were created and are shown in Figure 3.

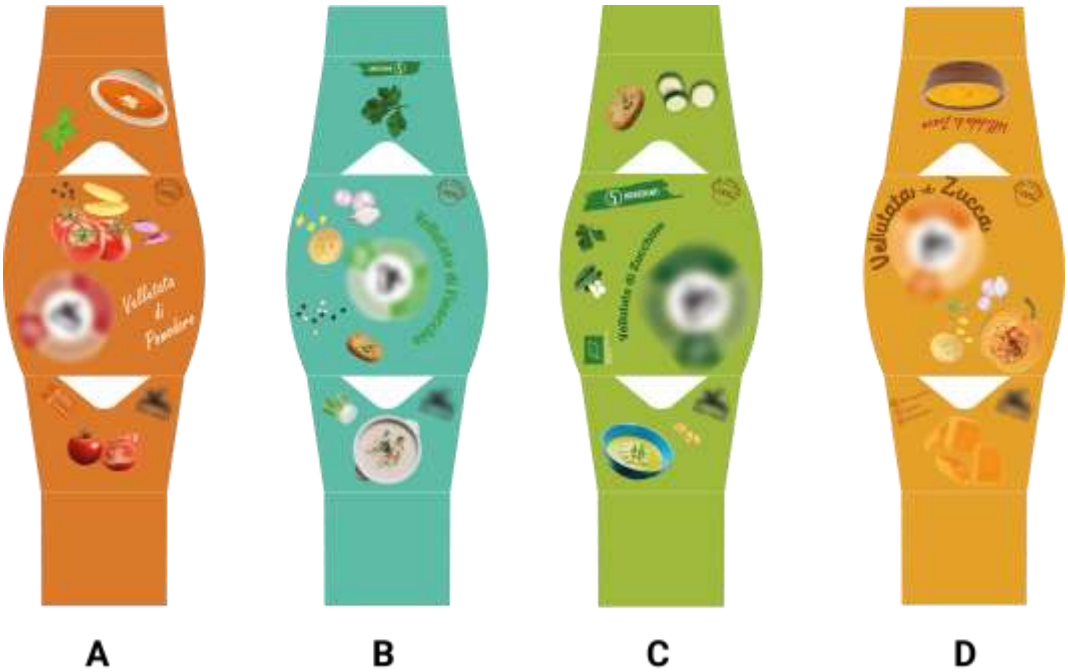

Figure 3: The four alternative layouts for the cardboard sleeve

At the beginning of each condition, participants were presented with one of the layouts printed on a cardboard and wrapped around the bowl as shown in Figure 4 and they were asked to accurately re-create the layout (i.e. position, rotation, scale of the assets, background colour) using the proposed

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

user interface. Each layout contains the same number of assets (12) and has the same distribution of assets across the three main faces (three on the front, seven on top, two on the back). All the assets are positioned, rotated and scaled arbitrarily except two items that are placed always at the same manner: on the top right corner of the top surface and on the top right of the front face. These assets are present in all the layout compositions, keeping the same position, rotation and scale as control properties.

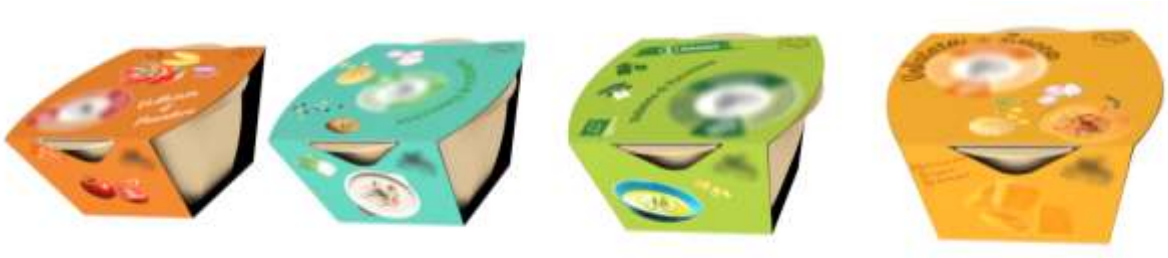

Figure 4: 3D view of the four alternative layouts

The experimental protocol is composed of several steps:

1. **Briefing:** Introduction and description of the test tasks to the user, signing of the participation and data recording consent form.
2. **Training for condition A:** Description of the user interface functionalities that will be used for the task A using a number of step-by-step videos.
3. **Free learning:** The user has few minutes to play with the user interface and learn about the functionalities of the interface.
4. **Condition A:** The user is informed that she/he has 10 minutes to reproduce one of the four layouts as accurately as possible, starting from a blank model. If the user requires more than 10 minutes, she/he can continue until the completion of the task. Once the user completes the task, she/he presses the save button in order to save the final layout configuration.
5. **Questionnaire 1:** After the first test, the user was asked to fill two subjective questionnaires to assess the usability of the system. The first one is the System Usability Scale (SUS) and the second one is the Creative Support Index (CSI), both are described in depth in section 4.4.2. The questionnaires have been provided to the user using Google Form (POLIMI and GINP) or in paper format (UBATH).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

6. **Training for condition B:** Since most of the main functions of the interface are the same of condition A, the training mainly focuses on the differences in terms of visualization and interaction modality, according to the B condition assigned to the user.
7. **Condition B:** The user has 10 minutes of time, as for condition A, to reproduce a different packaging layout using the tablet and the mixed prototype. In the case of condition B4 the user repeated the test using exactly the setup as they had used in condition A but were asked to replicate a different layout.
8. **Questionnaire 2:** After the second test the user was asked to fill again the two questionnaires (see point 5).

4.2 ORGANIZATION OF THE TEST

Please note that the content of this section is under temporary embargo until 31st January 2019 to allow for the filing of a patent application

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

Please note that the content of this section is under temporary embargo until 31st January 2019 to allow for the filing of a patent application

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.

H2020_ICT_2015_SPARK_688417

Please note that the content of this section is under temporary embargo until 31st January 2019 to allow for the filing of a patent application

4.3 PARTICIPANTS

The participants were all engineering students recruited from each of the three academic partners. The level of design knowledge and experienced varied but most participants were recruited from the later years of undergraduate courses or from masters/postgraduate courses and therefore were considered to have a good level of design knowledge. To further characterise the background of the participants, they were asked if they would describe themselves as a 'designer' or an 'engineer' and how many years of experience they had in using CAD software. The number of participants at each test location and their profile is summarised in Table 2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

TABLE 2. NUMBER OF PARTICIPANTS AND THEIR PROFILE FOR EACH TEST LOCATION

	UBATH	GINP	POLIMI
Number of participants	19	28	37
Number self-identifying as a 'designer'	7	-	8
Number of females	5	10	6
Average experience with CAD software (years)	6	-	5

4.4 DATA COLLECTION

The test included objective and subjective evaluation. A logging module has been expressly included within the software of the SPARK platform to monitor the students' activities. Whilst, specific questionnaires have been proposed to the students at the end of the test execution. This section describes the data collected within the log file and the questionnaires elaborated to evaluate the participants' perceptions of usability.

4.4.1 Log file

Different types of log data were collected during each condition of the test, in particular the modifications performed by the participants on the user interface (Activity Log) and the layout configuration at task completion (Saved Version). In conditions B (1, 2, 3), additional data regarding the position of the mixed prototype have been collected (Tracking Log). Below, all the types of log are described in detail:

- Activity Log – It captures all the 'events' initiated by the user through the user interface. The main types of event captured within this log were:
    - Selection and manipulation of assets (placement, rotation, re-scaling and layer order);
    - Activities related to use of the interface (asset filtering by tags, image swap and deletion);
    - Change of background colour;
    - Change of visualisation and change of viewpoint of the virtual model or UV Map.
- Saved Version – This log captured a snapshot of the system status, including the types of data captured in the Activity Log, when the 'screenshot' button was pressed within the user interface. The participant was asked to press this button when they had completed the task. This log was used to measure the task completion time and the data within this file was used to evaluate the accuracy of the position, rotation and scale of assets.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

- Tracking Log – The log generated by the Spatial Augmented Reality (SAR) module. This included:
    - Position (x, y, z coordinates) of the SAR model;
    - Rotation of the SAR model.

4.4.2 Participants' perceptions of usability

Two questionnaires were used to assess the participants' perception of the usability of the system, the Creativity Support Index (CSI) and the System Usability Scale (SUS). The CSI questionnaire has been developed based on the NASA TLX survey method, but with a greater emphasis on evaluating creativity support tools [4]. The CSI questionnaire was used within Task 4.2 and Task 5.1 of the SPARK project to gather feedback on designers' perception of the usability performance of SPARK. The survey features two sections. In the first section, the user rates the performance of the system they have tested in terms of its ability to support creativity factors, such as 'exploration' and 'immersion in the task'. The second section requires the user to determine which of these creativity factors are most relevant for the task they completed used the system. This enables the calculation of weightings for each of the creativity factors. Because of the very constrained nature of the design task (individuals working by themselves to copy an existing design), it was decided to modify the CSI survey to reflect the nature of the task. Two creativity factors were removed from the survey, namely the 'Expressiveness' and 'Collaboration' factors. The scoring calculations were adjusted to account for these modifications – see Section 5.2.1.

The SUS questionnaire is commonly used to assess the usability of products, software and websites [3]. An independent review of the reliability of the SUS tool based on 10 years of studies that have employed this questionnaire concluded that it is "a highly robust and versatile tool for usability professionals" [1]. The SUS questionnaire features 10 statements, including five positive statements (such as "I think that I would like to use this system frequently") and five negative statements (such as "I found the system unnecessarily complex"). The participant uses a five-point Likert scale to state their level of agreement with each of the statement. The scoring system produces a maximum score of 100. The CSI and SUS questionnaire templates are provided in Appendix A.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

4.5 TREATMENT OF THE DATA

4.5.1 Accuracy of placement, rotation and scaling of assets

The mean errors calculated with respect to the position, rotation and scale of assets were calculated using the 'Saved Version' data. For each accuracy metric, a number of steps were necessary to calculate the mean error for each condition. Here, we explain the process for the position error (the process for rotation and scale error were similar):

1. The target value was determined by taking the mean position of each asset for each layout. Using this mean value helped to reduce the impact on the results of any misalignment in the SAR projection, which improved the comparability of results from across the different test sites.
2. The Saved Version data for each trial (where a 'trial' refers to one participant completing one condition) was compared with the target values. The difference between the target value and the actual value was the error. The error in the x and y directions were combined into a total position error by calculating the length of the vector from the target position to the actual position.
3. The process was repeated for each trial and then the mean error was calculated for each condition.

For the rotation error metric, it was necessary to use unit vectors to compare the target and actual values to ensure that the smallest rotation error angle was calculated. For example, if the target value was 350° and the actual value was 5°, then the rotation error was calculated as 15° (and not 335°).

4.5.2 Accuracy of colour

The evaluation of the accuracy of the colour selection mainly relates to the personal sensitivity, which the user has in comparing the colour projected onto the mixed prototype with to the one of the real prototype. To evaluate the difference between the two colours, the *CIEDE2000* colour-difference formula has been used [5]. This formula allows evaluating the colour difference in the *CIE L*C*h** colour space where three components define a specific colour. The three components are:

- *L**: lightness, where 0 means black and 100 is the maximum light intensity which is still visible without causing eye damage;

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

- *C**: Chroma, where 0 means completely unsaturated colour (i.e. a neutral grey, black or white) and 100 the maximum "colour purity";
- *H**: hue, considering the colour shades distributed in a circle, the units are in the form of degrees (or angles), ranging from 0° (red) through 90° (yellow), 180° (green), 270° (blue) and back to 0°.

The *CIEDE2000* formula combines the three colour components and provides a unique output named ΔE, which represents the overall difference between two similar colours. Whether the output is lower than the *Just Noticeable Difference* (*JND*) threshold, the two colours can be considered equal. Since during the colour calibration procedure of the prototype we accepted a colour disparity around 2.3 we can consider 5 a good *JND* value for the evaluation.

4.5.3 Efficiency

The measure of efficiency used was the time taken to complete the task. At the beginning of the task, the 'screenshot' button was pressed to create a timestamp and the same was done when the participant said that he/she had completed the task. The time elapsed between the timestamps of these two Saved Version files was calculated and used as the official duration of the task.

4.5.4 Usability

The SUS survey data were scored using the conventional scoring scheme, which generates a total score out of 100.

For the CSI survey, two factors had been excluded from the survey ('Collaboration' and 'Expressiveness') and so it was necessary to adjust the scoring scheme. The maximum CSI raw score is 300 when using the standard CSI survey with six factors. This raw score is normally divided by a factor of three so that the final score is a mark out of 100. The modified CSI included four factors, meaning that the maximum raw score was reduced to 120. Therefore, in order to keep the final score as a mark out of 100, the raw score was divided by a factor of 1.2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

5 RESULTS

5.1 ACCURACY

The following section reports the accuracy results for position, rotation and scale of the assets. The data collected from the two academic partners POLIMI and UBATH have been treated according to the process described in section 4.5. The data collected at GINP partially referred to a layout with an inverted configuration and, as such, they are not directly comparable with each other. Therefore, they will be further processed, to complete the comparative analysis with all collected data.

It was also possible to compare the performance of the interaction modalities in the different experimental setup between POLIMI and UBATH (i.e. tracking system and projectors configurations illustrated in section 4.2). The final section (5.1.4) covers the results of the colour accuracy tests carried out at POLIMI, as outlined in 4.5.2.

5.1.1 Accuracy of placement

POLIMI results for placement accuracy (Figure 8) show that there are not substantial differences in between the condition A with mouse input and the conditions B (1, 2, 3) with the multi-touch tablet interface. The mean accuracy values and the dispersion of data are consistent with each condition. Among the B conditions, the 3D view (B3) has a more compact and less disperse data set, even if the other conditions B (1, 2) are not very far in terms of dispersion and placement accuracy.

Condition B4 resulted less accurate than condition A and it shows that there is no significant learning effect in terms of placement accuracy of assets, at least in such a short term. This could also be due to the lower user attention in repeating the same task with the same interface.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

UBATH results for placement accuracy (Figure 8) share the same trend of the POLIMI data. Condition A has most of the results below the 2 mm, scoring slightly better than POLIMI while in B4 there is a higher dispersion of results. In conditions B2 and B3 the accuracy is higher than POLIMI, in particular with the Touch Area there is a dense concentration of results with very high accuracy.

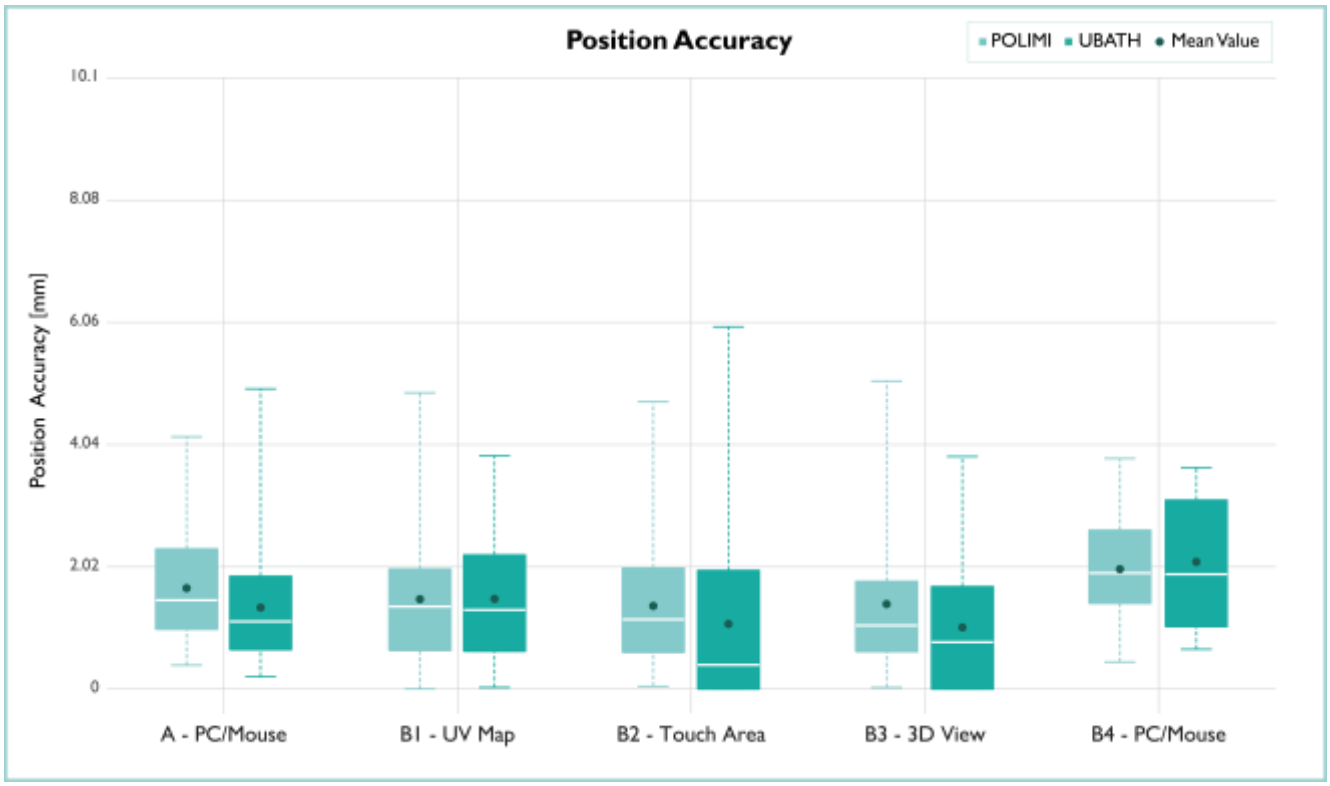

Figure 5: Box plot illustrating the position accuracy [mm] of the assets related to each experimental condition and the setup of the academic partners POLIMI and UBATH

5.1.2 Accuracy of rotation

POLIMI rotation accuracy results for conditions A and B4 are similar between them. In B4 there are no improvements in terms of rotation accuracy, therefore also in this case there is no evidence of any learning effect. Conditions B1 and B2 have also consistent results with little advantage for B2 where the dispersion of the results tends to be higher in the second quartile. B3 has more dispersed data than conditions B1 and B2 therefore, for the rotation function, the touch area and the UV Map visualization scored better.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

UBATH results for conditions A and B4 show a wider data dispersion and less accuracy by looking the mean values. In particular, in B4 some results bring the upper adjacent high in the diagram with errors greater than 35°. Condition B1 and B3 scored similarly between each other and generally better than B2. This could be due to the current limitations in the movement and handling of the mixed prototype using the rotary tracking system that makes the Touch Area interaction modality more difficult to use.

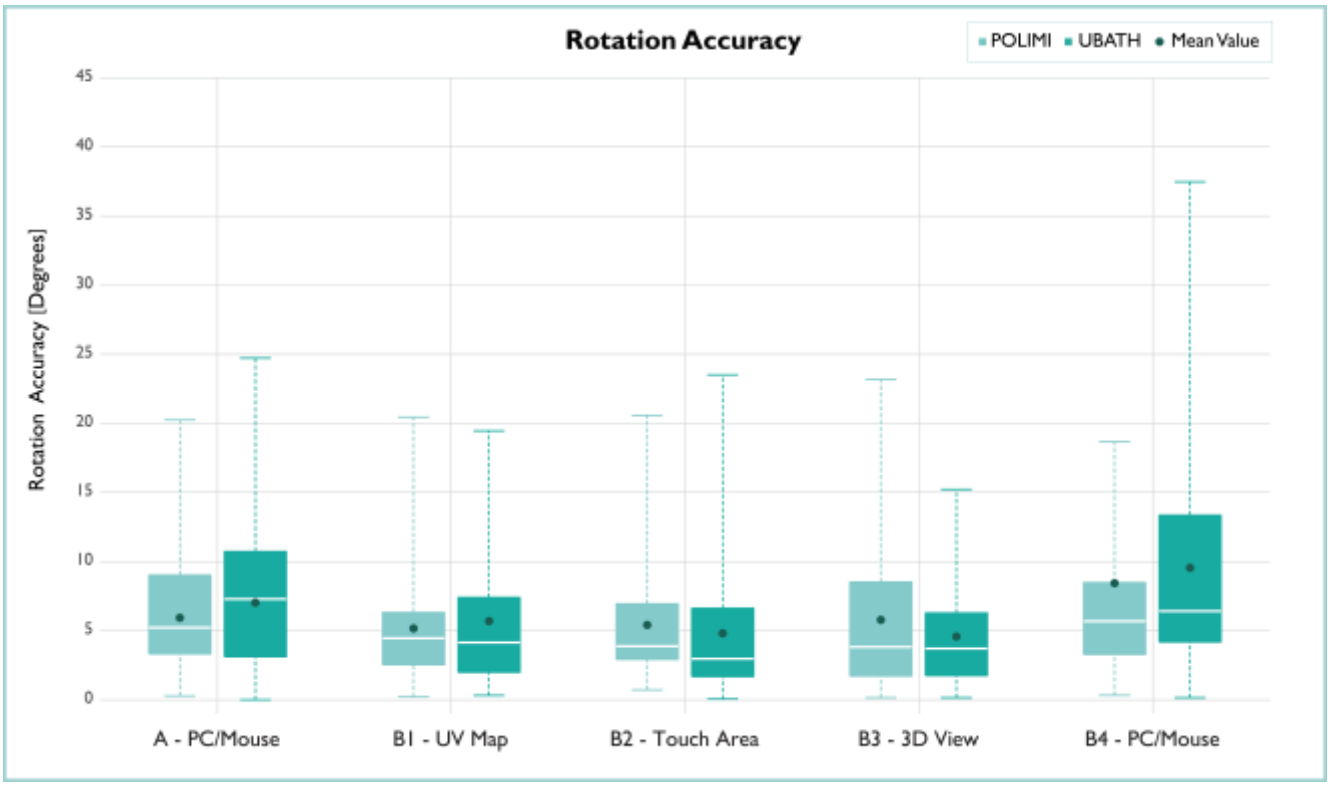

Figure 6: Box plot illustrating the rotation accuracy [Degrees] of the assets related to each experimental condition and the setup of the academic partners POLIMI and UBATH

5.1.3 Accuracy of scaling

POLIMI results (Figure 10) with condition A highlight better scale accuracy than with condition B4 that shows more dispersed results and greater upper and lower limits. In addition, still no visible learning effect appears after the first round with condition A.

Condition B1 shows a higher density in the first quartile with respect to condition B2 and it is showing scores that are more consistent. Condition B3 emerged to be the less accurate among conditions B (1,2,3) having a wider spread of the data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

UBATH results (Figure 10) for conditions A and B4 have similar results with POLIMI, A seems to be more accurate and there are no learning effects visible. Conditions B (1, 2, 3) have in general wider dispersion of data than POLIMI results. In particular, condition B3 has more consistent and focused results. Condition B2 also in this case scored worst highlighting, as in the rotation accuracy, the possible limitation due to the rotary tracking system.

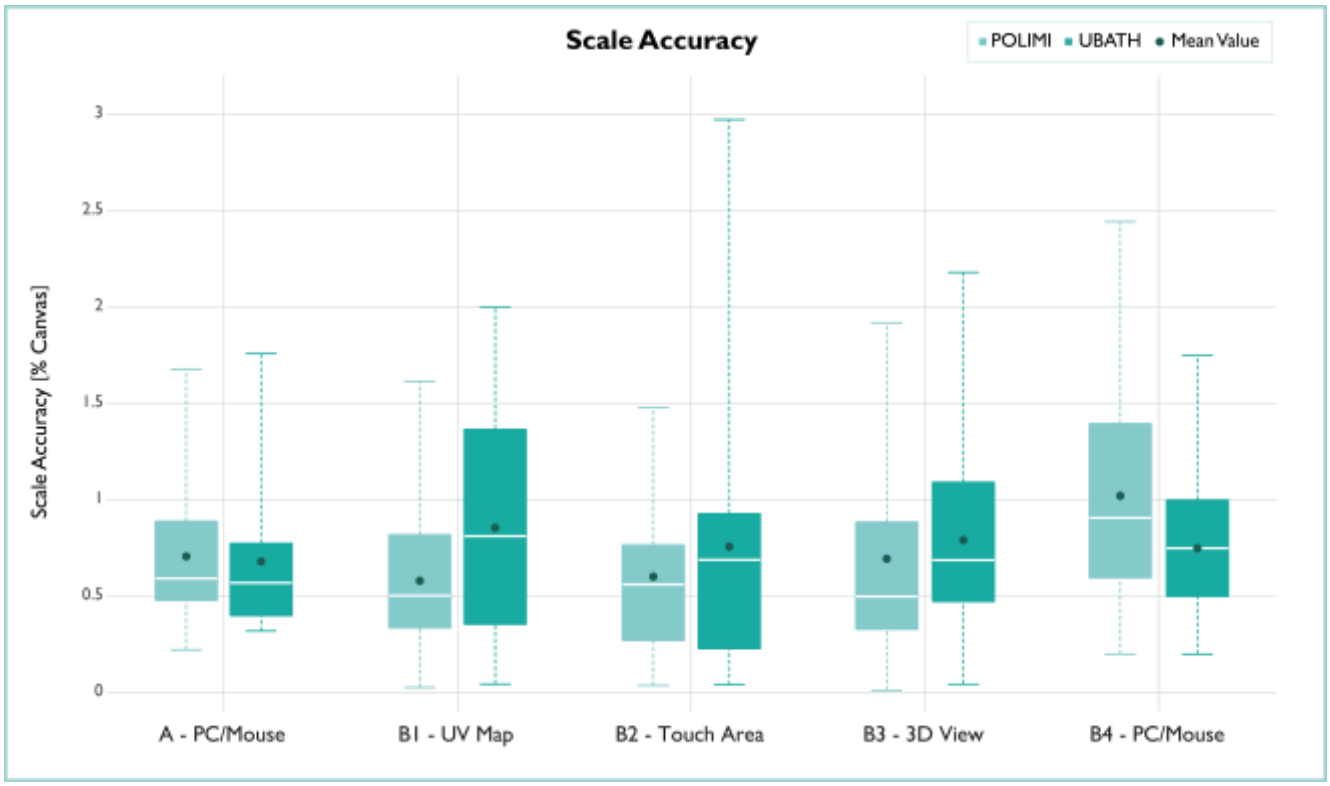

Figure 7: Box plot illustrating the scale accuracy [%Canvas] of the assets related to each experimental condition and the setup of the academic partners POLIMI and UBATH. Canvas is the 2D space that groups all the assets of the layout. Every asset placed inside the canvas is rendered in real time on the 3D model.

5.1.4 Accuracy of colour

The evaluation of the colour accuracy has been conducted on the users involved in tests of the interface B1, B2 and B3 at POLIMI, where an accurate colour calibration of the mixed prototype has been performed. The users involved in A and B4 have been excluded since they used the monitor as means for colour evaluation.

Figure 11shows the mean and standard deviation of ΔE calculated according to the colour, which the users had to select during the test. The value of the ΔE is very high, especially for the green colour (prototype C), but, if we analyse the contribution of the three colour components, ΔE mostly depends on the error related to the component L. Chroma and hue values, instead, seem to be very close to

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

the real colour. If we check the value of ΔE, calculated without considering the effect of lightness (ΔECH), in fact, it is drastically lower. We should conclude that the projected colour is perceived very close to the real one in terms Chroma and hue while the lightness component is not so easy to evaluate, as it emerges by the average value and standard deviation of $|\Delta L|$. In general, users perceive projected colour brighter than the reference colour and this is due to the intrinsic brightness of projected images with respect to the real colour that is not illuminated by the same light intensity.

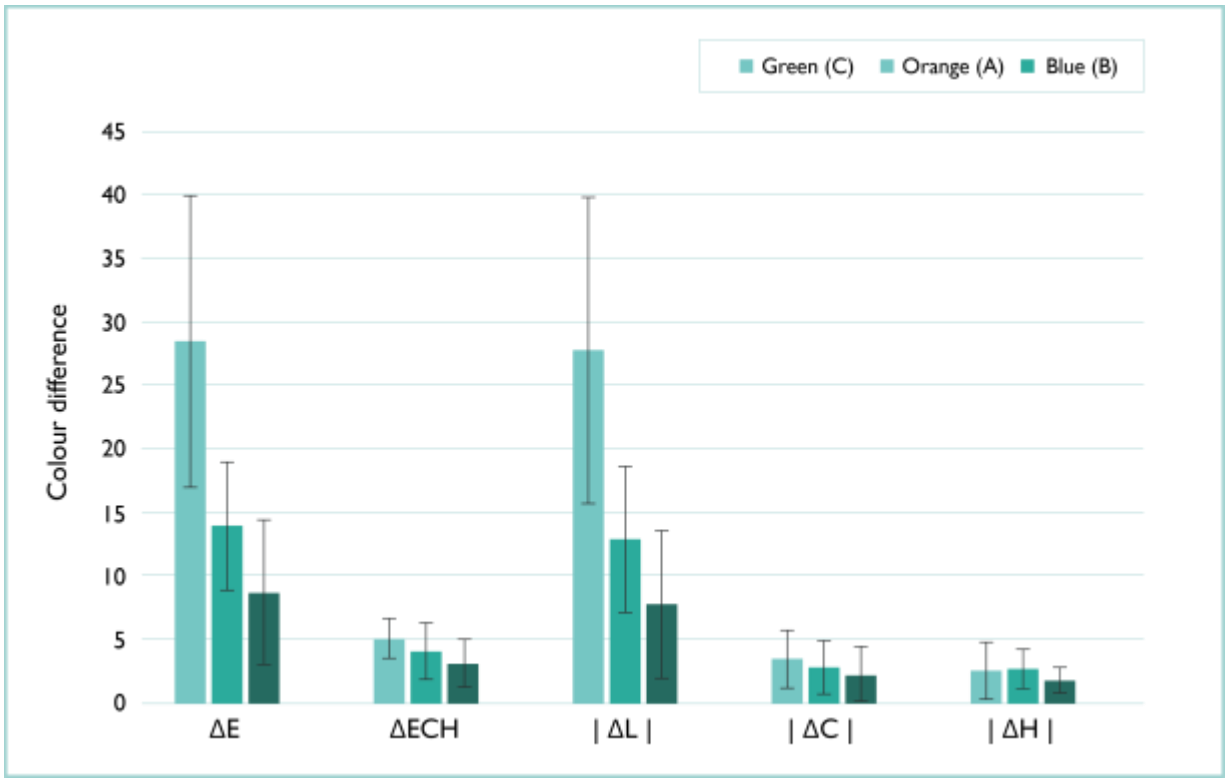

Figure 8: Chart related to the mean values and standard deviation of the CIEDE2000 [] colour difference (ΔE), CIEDE2000 [] colour difference calculated without the ΔL component (ΔECH), lightness difference (|ΔL|), Chroma differences (|ΔC|), hue difference (|ΔH|)

5.2 EFFICIENCY

The analysis of the efficiency shows a significant reduction of the completion time between condition A and B4, the two executed with the mouse as the mean of interaction, for both the universities. This trend can be explained by taking in consideration the learning effect of the interface and interaction (the users during the second test has a higher self-confidence and a clearer idea of how a modification should be performed on the UI) and the repetitiveness of the task (the user is less engaged to complete each modification with the same accuracy).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

Higher disagreement was met instead in the comparison of the results of the two universities for the tablet conditions (B1, B2 and B3). All the three conditions for POLIMI have a mean value higher than the threshold of 10 minutes while for UBATH all the data are below the established limit (with the only exception of one user in condition B2). This is a consequence of the different setup adopted (Section 4.2) at UBATH where the limited movement of the augmented prototype has forced the user to be focused mainly on the virtual contents displayed on the tablet. In POLIMI instead, since all the users were asked to consider as reference output the mixed (SAR) prototype, there is an increased difficulty of the task perceived by the tester which had to control simultaneously the physical target layout (the printed one) the digital contents (available on the tablet) and the augmented prototype (Figure 12).

This latter aspect is highlighted in Figure 13 where the mean time for each condition is partitioned between four different activities:

- the time spent to identify the correct background colour;
- the time spent to manipulate the 3D model/UV map on the tablet;
- the time spent to manipulate the tracked prototype;

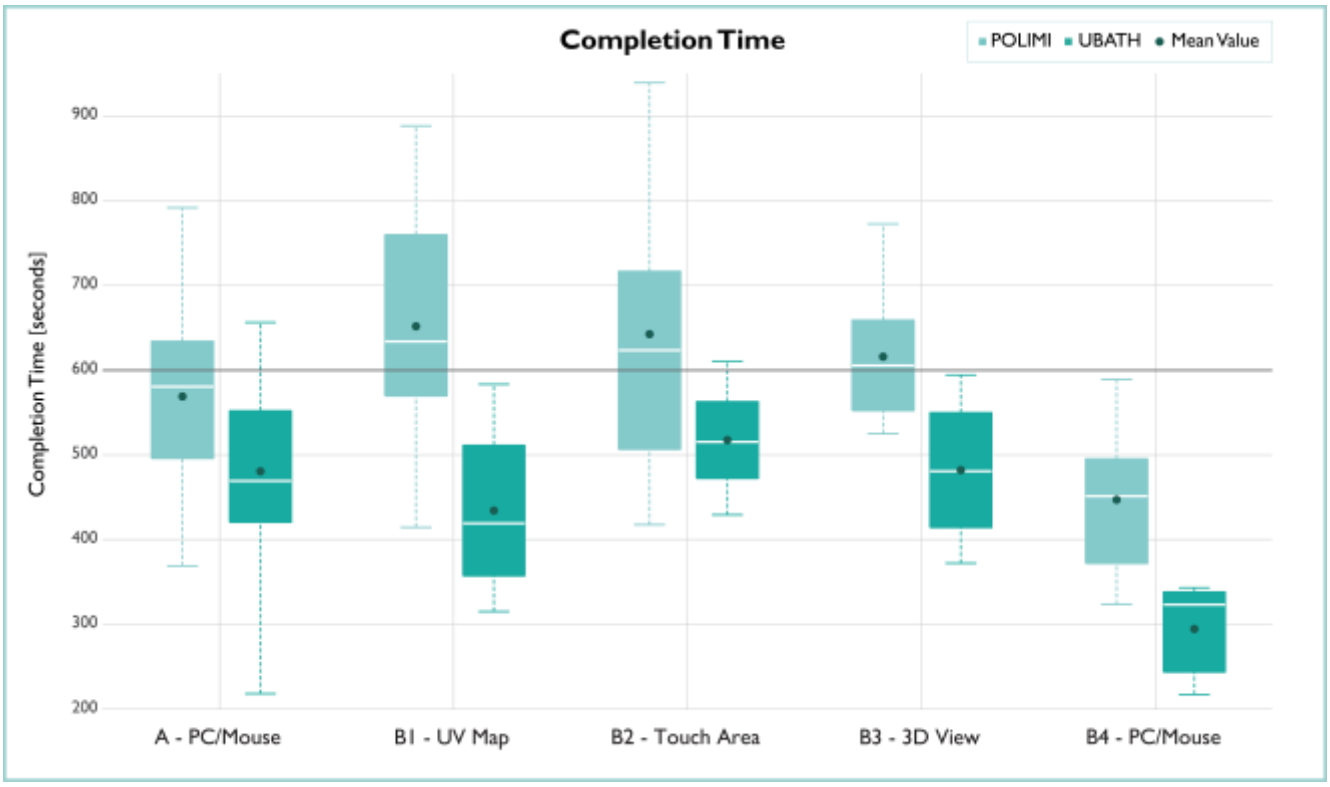

Figure 9: Box plot illustrating the completion time [seconds] related to each experimental condition and the setup of the academic partners POLIMI and UBATH

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

- the time spent to interact with the UI or to look at the target layout.

Especially for conditions B1, B2 and B3, the users at POLIMI invest more than 10% of their available time to perform the first three "side" activities while at UBATH this percentage drop below 5%. The only exception is for condition B3 where most of the user had encountered high difficulties to orbit the camera of the virtual model.

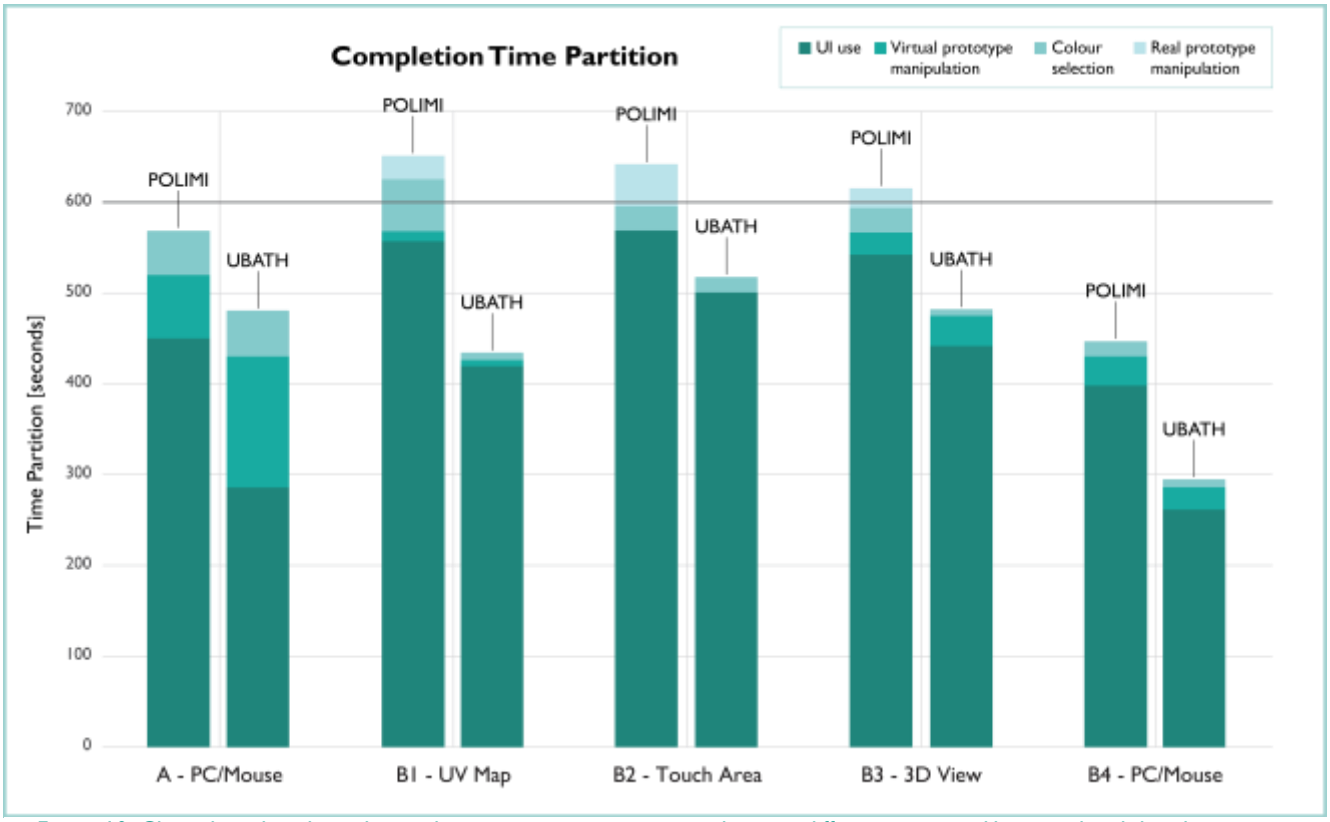

Figure 10: Chart describing how the total execution time is partitioned among different activities like virtual and digital prototype manipulation, colour selection and UI interaction for each experimental condition and setup of the academic partners POLIMI and UBATH.

5.3 PARTICIPANTS' PERCEPTIONS OF USABILITY

The participants' perceptions of the usability of the various interfaces were assessed using the System Usability Scale (SUS) and the Creativity Support Index (CSI) surveys. Considering first the results of the SUS survey, a total of 160 responses were collected, including 74 from POLIMI, 43 from GINP, and 43 from UBATH. The SUS mean scores for each condition from each of the partners are presented in Figure 14 along with an adjective grade ('Excellent', 'Good' and 'OK') based on the recommendations of [2].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

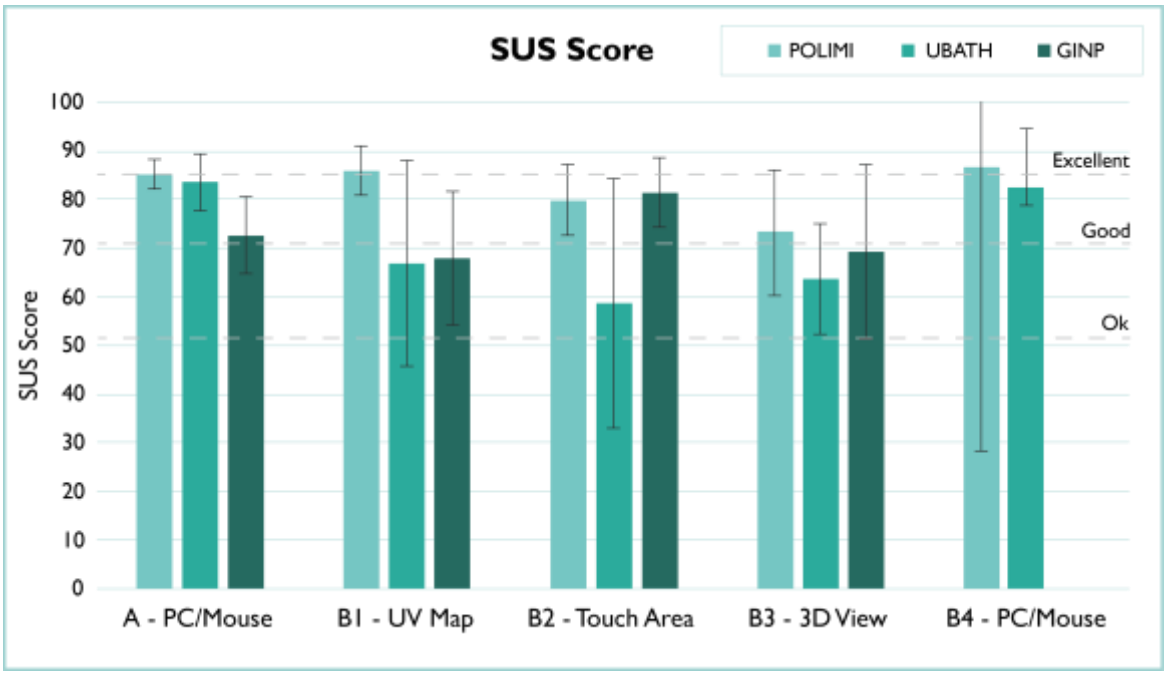

Figure 11: SUS mean score with confidence interval shown

It is important to note that the confidence interval is large for some of the results, notably the UBATH results for all the 'B' conditions, and the GINP data for conditions B1 and B3. Hence, where there are discrepancies in the results between the three test locations, more emphasis has been given to the results that show a smaller confidence interval (higher confidence).

The first observation is that conditions A and B4, which both featured the PC screen and mouse interface, scored consistently highly in the 'Good' to 'Excellent' range (no GINP data available for condition B4). This is to be expected given that all participants had experience of using CAD software, which makes use of this type of typical PC/mouse interface. The similarity in the scores between conditions A and B4 suggests that there was no significant learning effect.

If we ignore the B1 and B2 condition results from UBATH due to the large variability of the results, the worst performing interface was the 3D view and tablet used in condition B3. Whilst the scores were in the 'OK' to 'Good' range, it was surprising that this was the worst performing condition given that the visualisation within the user interface is identical to that used in conditions A and B4. One possible explanation was that during the experiments, the researchers observed that several participants seemed to struggle with 'orbiting' the 3D virtual model to obtain their desired viewpoint.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

In particular, the orbit function requires multiple swipes across the screen to change the view model from the front face to the back face.

Conditions B1 and B2 both performed well. For condition B1 there is a reasonable large discrepancy between the mean score from POLIMI (mean=86) and GINP (mean=68), although more weight should be given to the POLIMI result due to the smaller confidence interval. The results for condition B2 are much more consistent between POLIMI (mean=80) and GINP (mean=81.4). The much lower mean score for condition B2 from the UBATH results (mean=58.8) should be ignored due to the very large confidence interval and the likely influence of the difference in tracking technology between the rotational tracking system used at UBATH and the full tracking systems used at GINP and POLIMI.
For the CSI survey, only UBATH and POLIMI collected data as the initial results collected showed very similar patterns to the SUS survey data. This allowed GINP to skip this part of the protocol, freeing up time for them to include more participants.
From Figure 15 it can be seen that the scoring patterns from the SUS results are largely repeated with the CSI results. Conditions A and B4 (PC/mouse) performed best followed by condition B1 (UV Map). A difference, instead, emerged with the CSI scores of POLIMI where the condition B2 (Touch Area) is better than B3 (3D View), conversely to what emerged in the SUS analysis.

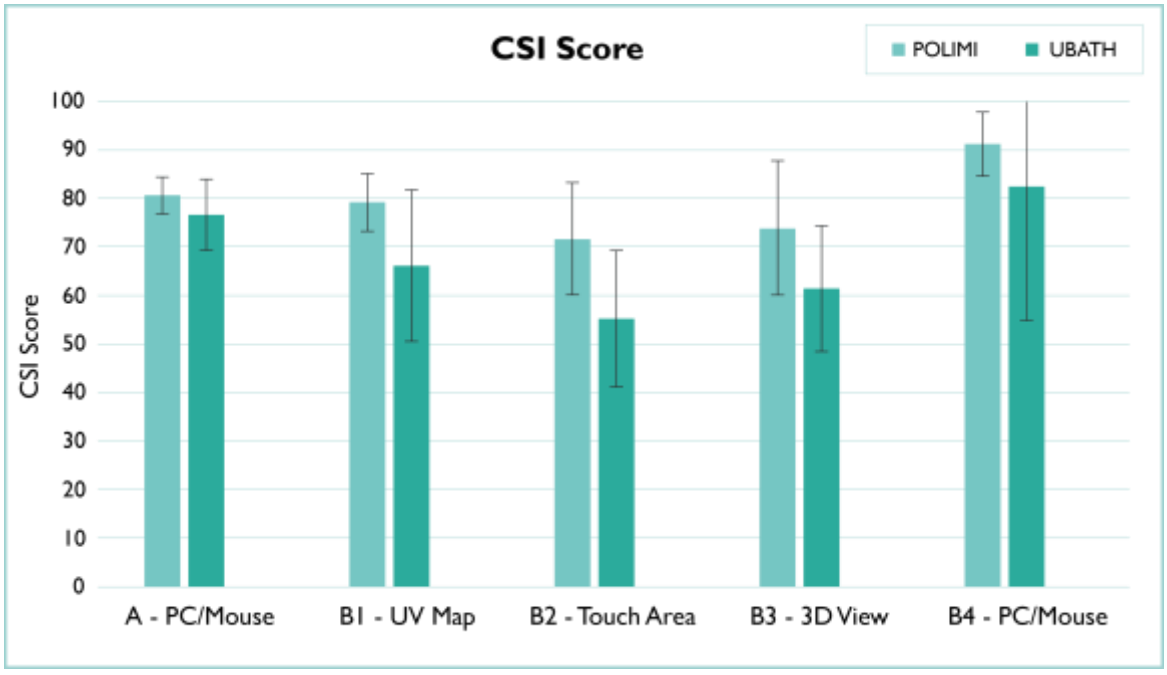

Figure 12: CSI mean score with confidence interval shown

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

The overall conclusions from the analysis of the usability data are:

- Conditions A and B4, which both made use of the PC/mouse interface, were consistently higher than all other conditions.
- Of the tablet-based interfaces, the UV Map (condition B1) performed the best.
- The worst performing interface according to the SUS results was the 3D View (condition B3) – this may be due to difficulties experienced by participants in orbiting the virtual 3D model using the tablet interface.
- The SUS and CSI survey produced similar patterns of results and so it is recommended that only the SUS survey be used in the future to save time and effort.

6 CONCLUSIONS

In this document, we presented the results of the testing activities carried out with a total of 84 students of the three academic partners involved in the project: POLIMI, GINP and UBATH.

The main objective of these tests was to validate different interaction modalities, based on multi-touch gestures on a tablet screen, to manipulate the digital contents projected on the augmented prototype of the SPARK platform. This validation has been performed taking in consideration the usability perception, which was assessed through CSI and SUS questionnaires, the efficiency, in terms of task-completion time, and the accuracy in assets and colour manipulation. These aspects have been further compared with the most common-used interfaces (i.e. mouse and keyboard).

Relevant outcomes, with reference to the hypothesis of Section 3.2, can be summarised as follows:

- The accuracy, in terms of position, rotation and scale, measured within the touch conditions (B1, B2, B3) is comparable with the mouse interface (supposed to be the most accurate due to the greater confidence for a common user) and, in some cases, it has even a lower error rate. Thanks to that, we can assume also the positive impact of the SAR environment in comparing the rightness of an asset located on a real or on a mixed prototype.
- Different tracking systems and graphical render quality influence the accuracy of the tasks completed with the tablet. Increasing the manipulation possibilities of the augmented prototype allows the users to have a better understanding of how they are performing the task.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

- Within the same setup, there is no relevant difference in term of accuracy between conditions B1, B2 and B3.
- The task related to the colour selection revealed that, with the SPARK platform, the users were able to select colours very close to the reference, which demonstrates the effectiveness of the colour calibration procedure applied to the projectors. In addition, the time needed to make the selection is comparable with the one spent by using the monitor as visualisation device.
- The efficiency has a significant variation between different conditions. At POLIMI touch condition has been more time consuming than the mouse interface (higher than the limit of 10 minutes) due to the increased complexity of the task for the introduction of the mixed prototype as target. At UBATH, instead, there is no a significant trend even if, in most of the cases the, B condition has been completed faster than the A.
- All the users, independently from the setup, have perceived a better usability of the mouse conditions in comparison with the touch interfaces. This is justified by the greater experience they have with this kind of interaction device.
- The B1 condition (touch with UV map view) has the best-perceived usability among all the others tablet conditions. This could be due to the fact that all the assets of the layout are visible in one single view (i.e. no need to continuously rotate a 3D model) and the outline of the packaging adds more helpful information for the user when placing the assets;
- A cross analysis between efficiency and accuracy has suggested that the condition B4 has been performed with a low-level participation of the user due to the repetitiveness of the task; while the high precision achieved by the touch conditions required more time to the user;
- The learning effect does not influence the measurement of the accuracy and the usability, but it is relevant for the efficiency.

Therefore, we can conclude that also these tests with students confirm the good level of usability reached by the SPARK platform, which can be compared with much more consolidated systems.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

7 REFERENCES

[1] Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, *24*(6), 574-594.

[2] Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114-123.

[3] Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

[4] Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *21*(4), 21.

[5] Sharma, G., Wu, W. and Dalal, E. N. (2005), The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. Color Res. Appl., 30: 21-30. doi:10.1002/col.20070

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.

H2020_ICT_2015_SPARK_688417

8 APPENDIX A

Creativity Support Index survey

Name: **ICT technology used:**

Please rate your agreement with the following statements:

I was satisfied with what I got out of the system or tool.
Highly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **Highly Agree**

It was easy for me to explore many different ideas, options, designs or outcomes, using this system or tool.
Highly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **Highly Agree**

I would be happy to use this system or tool on a regular basis.
Highly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **Highly Agree**

My attention was fully tuned to the activity, and I forgot about the system or tool that I was using.
Highly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **Highly Agree**

I enjoyed using this system or tool.
Highly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **Highly Agree**

The system or tool was helpful in allowing me to track different ideas, outcomes or possibilities.
Highly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **Highly Agree**

What I was able to produce was worth the effort I had to exert to produce it.
Highly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **Highly Agree**

I became so absorbed in the activity that I forgot about the system or tool that I was using.
Highly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **Highly Agree**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

Section 2.

When doing this task, it's most important that I'm able to…

Explore many different ideas, outcomes, or possibilities	☐	☐	Work with other people
Be creative and expressive	☐	☐	Produce results that are worth the effort I put in
Enjoy using the system or tool	☐	☐	Become immersed in the activity
Become immersed in the activity	☐	☐	Produce results that are worth the effort I put in
Work with other people	☐	☐	Enjoy using the system or tool
Produce results that are worth the effort I put in	☐	☐	Explore many different ideas, outcomes, or possibilities
Be creative and expressive	☐	☐	Become immersed in the activity
Work with other people	☐	☐	Produce results that are worth the effort I put in
Be creative and expressive	☐	☐	Enjoy using the system or tool
Explore many different ideas, outcomes, or possibilities	☐	☐	Become immersed in the activity
Work with other people	☐	☐	Be creative and expressive
Produce results that are worth the effort I put in	☐	☐	Enjoy using the system or tool
Explore many different ideas, outcomes, or possibilities	☐	☐	Be creative and expressive
Work with other people	☐	☐	Become immersed in the activity
Explore many different ideas, outcomes, or possibilities	☐	☐	Enjoy using the system or tool

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417

System Usability Scale survey

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.688417.
H2020_ICT_2015_SPARK_688417