Project Number	621668-EPP-1-2020-1-ES-EPPKA2-KA
Project Title	Mixed Reality in medical Education based on Interactive Applications (MIREIA)

D7.1. VALIDATION PLAN

Dissemination Level	Public
Delivery Date	30/06/2022
Responsible	CCMIJU
Authors	Juan A. Sánchez-Margallo, Francisco M. Sánchez-Margallo, Èlia Pedregosa Sandoval, Nikos Skarmeas, Thomas Langø, Calin Tiu, Ignacio Oropesa García, Carmen Guzmán García, Enrique J. Gómez Aguilera, Patricia Sánchez González, Ole Jakob Elle, Magdalena K. Chmarra, Jenny Dankelman

Co-funded by the Erasmus+ Programme of the European Union

Version control

Version	Date	Comment
01	09/06/2022	First version of the report (CCMIJU)
02	24/06/2022	UPM updates
03	28/06/2022	MEDIS and AVACA updates
04	30/06/2022	CCMIJU, SINTEF and OUS updates
05	05/07/2022	TUDELFT updates
06	16/08/2022	CCMIJU updates
07	28/02/2023	CCMIJU updates

1. Summary

This document defines and explains the validation roadmap both technical and clinical (subjective and objective) with regard to the tools and contents developed in the previous work packages. It will bring together all the agreements reached by the partners on the validation methodology: number of participants, type of trials, tools used for validation, statistical and/or qualitative analyses to be performed or any aspect that may be necessary to reach an agreement on each particular trial. As the validation process strongly depends on various project results, this document (deliverable D7.1) could be subject to possible updates. Both task T7.1 and deliverable D7.1 are coordinated by CCMIJU.

Table of contents

1. Introduction

Deliverable D7.1 describes the validation plan to be followed during the validation phase of the MIREIA project results, including its repository platform and the various types of educational content offered. The validation plan is based on the expertise of the Consortium members, and includes the description of the methodology to be followed at each stage of the validation process.

Since the results of the MIREIA project are mainly technological, we will organize the validation roadmap into: (1) technological verification of proposed solution; (2) subjective validation of the MIREA solution; and (3) objective validation of the MIREIA results. The technological verification will be conducted firstly, followed by the validation of the MIREIA results. Finally, an analysis phase of the data obtained from these evaluations will be carried out in order to draw the pertinent conclusions.

The terms *verification* and *validation* are defined as follows:

1.1. Verification

Verification is the process of checking whether a software achieves its goal without any bugs (code problem). This is the process to ensure whether the software that is developed is working right or not. It verifies whether the developed software fulfils the requirements that we have for the correct functioning according to the D4.1. Specifications.

Verification is a **Static Testing**. Activities involved in verification are:

1. Inspection
2. Review
3. Walkthrough
4. Desk-checking

And these activities span to user and system requirements, to all levels of design, and system verification.

1.2. Validation

Validation is the process to demonstrate the effectiveness of a learning tool. Therefore, validation is an essential step when proposing new forms of learning [Colardyn2009]. Different methodologies should be considered to cover the various aspects under validation and to extract measures (objective and subjective) to identify weaknesses and strengths of these learning tools. Therefore, validation is a **Dynamic Testing** which follows the verification process. For this project, both subjective and objective validations will be performed with regard to the MIREIA results presented in WP5 (Implementation and Integration) and WP6 (Creation of Learning Contents). For these validations, end-users will be recruited from each clinical partner (CCMIJU, St. Olavs Hospital, Oslo University Hospital and MEDIS Foundation).

The different phases of the validation plan depend heavily on the results generated in the previous work packages, such as: the pedagogical needs identified and the methodological guidelines defined (WP3), the technical and functional requirements of the educational tools and contents to

be developed (WP4), and the implementation and integration of these tools and contents (WP5 and WP6).

2. Technological verification

During the technological verification, the functional requirements and specifications of the developed technological solutions will be assessed. These technical and functional specifications are defined in the deliverable **D4.1. Specifications**.

Table 1 shows the application components of MIREIA that need to be verified.

Table 1. List of technological solutions to be verified.

Application	SubArea	Responsible
Content repository	Storage and management of contents	AVACA
Content repository	Administration & User Management	AVACA
Content repository	Content Presentation & eLearning Platform	AVACA
Immersive visualization technology	Visualization of 3D models in the form of holograms	OUS
Semi-automatic creation of 3D models	Creation of 3D models by means of endoscopic video feed	eCapture3D
System for creating virtual environments for MIS training	Platform to design virtual environments for Minimally Invasive Surgery (MIS) training	UPM

During the technological verification, possible bugs (code malfunctions), improvements, and possible points of weakness of the technological solutions proposed in the project should be discovered. That is needed in order to ensure that the technological tools meet the expectations and increase the functional reliability of the system.

Technological verification will be led by experts of AVACA in conjunction with the support of the development teams of each responsible party described in the Table 1 (OUS, eCapture3D and UPM).

The results of this verification will have a clear effect on the implementation and integration work package (WP5), which may lead to the updating of some of the tools or technological developments. These possible updates will be done in order to solve possible problems detected or improvements to be taken into account.

2.1. Tools to be used for the verification

Technological verification will be done using **checklists and reviews** (see Table 2 for the whole list and explanation of the tools used during verification process). These checklists and reviews should assess the positive interaction and functioning of the technological solutions according to the technical and functional specifications (D4.1. Specifications).

Depending on the type of testing (ex unit testing, system testing etc) and the complexity of the scenarios at hand, testing is completed in automated or manual way. **Automated** testing/verification is done using automated scripts which simulate some scenarios and are executed via particular tools. An example of this is the definition of web service calls in an (input, expected output) manner and tools like Postman are used which by getting the input, execute the call and return the output for checking without the need of the front-end (UI) of the application, **Manual** testing/verification is used in more complex scenarios where there are steps with information need human effort to be generated (ex generation of a 3D models with input needed by human for defining various qualitative parameters or actions. Also, there are phases like specification etc which are by nature done purely by human effort.

Table 2. List of tools and tests for technological verification.

##	Activity	Type	Auto/Manual	Tool	Approach
1	Requirements Specification	Verification	**Manual**	Review	**Review** of requirements and if they can be met by the system design
2	High Level Design	Verification	**Manual** review of design documents to verify whether they offer the appropriate functionality	Checklist	A **checklist** of initial user requirements in scope of the application and if they need each relevant item.
3	Detailed Design	Verification	**Manual**	Review	**Review** of detailed design documents to verify they meet proposal prerequisites and Constrings.
4	Program Specification	Verification	**Manual**, partially out of scope	Checklist	Based on the integration scenarios provided during API design and the actual API Specs that the systems should offer, a **checklist** of APIs and (short) description of how they meet them should be produced
5	Coding	Both verification and	Auto	SonarQube	Code Checking. This involves static analysis which highlights potential

		validation			sources of errors, bad quality code etc. SonarQube is one of the most popular tools, it is also open source and to be used during the development processes. The above involves code Verification Scans and reports
6	Unit Testing	Validation	Automated and/or manual	**Postman** for testing Web Service scenations/	Unit Testing activities. This involves **Test Cases** which are highlighted during the definition of the user stories are tested,
7	Integration Testing	Validation	Automated and/or manual	**Automated scripts** for testing integration scenarios (both success and failure conditions) making use of **Postman** for testing Web Service scenations	Integration Scenarios covering both Business (content upload and view) and System Requirements (user registration and authentication) Integration scenarios are both manually and automatically are tested depending on the actual use cases. Auto testing involves rest service-based testing with autonated scripts with the use of postman. Manual testing is used (more hybrid actually) with multiple step scenarios with both manual steps where this is required and the rest as auto steps. Both success and failure scenarios including edge cases.
8	System Testing	Validation	Manual	**User Stories** and **Test Cases which**	Manual Testing to find out whether the relevant cases are met.
9	User Acceptance Testing	Validation	Manual	User Sessions	User session and meetings to verify user acceptance of the application

The technological verification will be conducted during the M21 (September 2022) and M22 (October 2022). A report of this process will be provided by M22 (October 2022).

3. Subjective validation

A subjective validation of the developed results will be conducted at CCMIJU, MEDIS, St. Olavs hospital and OUS. Each participant will take part in case studies related to anatomy, laparoscopy and flexible endoscopy, with a total number of users as specified in Table 3. The same user could participate in different case studies if it is considered appropriate, seeking the optimization of the user participation.

Table 3. Minimum number of participants per institution and use case for the subjective validation.

Use case / Partner	Anatomy	Laparoscopy	Flexible endoscopy
CCMIJU	30	30	30
MEDIS	30	30	30
St. Olavs	30	30	30
OUS	30	30	30

The MIREIA platform will be subjectively evaluated as an educational tool for:

1. Learning medical anatomy
2. Learning laparoscopy
3. Learning flexible endoscopy

These are the assessments to be conducted during this subjective validation:

- **Content validation**: It determines the degree of appropriateness of the material to the learning purpose (medical anatomy, laparoscopy and/or flexible endoscopy).
- **Usability validation**: It assesses the comprehensiveness, operativity and attractiveness of the learning environment (MIREIA learning platform).
- **Functionality validation**: It assesses aspects related to the functionality of the MIREIA platform.

The main aim of the validation is to discover possible problems, and improvements of the developed tools and contents (results from WP5 and WP6) before broadening the scope and reaching out to external end-users (e.g., residents), in order to point out the strengths and weaknesses of the proposed MIREIA solution. This will provide insights into how the platform will be used by real end-users. Conversely, it will help said end-users gain experience with the MIREIA platform before its launch and, this way, their valuable feedback can effectively be implemented to increase the usability of the MIREIA platform. Finally, it will help to improve and redefine the final cases as well as determine potential adjustments.

3.1. Content validation

This validation determines the appropriateness of the educational tools and materials provided for learning purposes in the field of medical anatomy, laparoscopy and/or flexible endoscopy. We

will use questionnaires for this assessment to get the first insights into the validity of the project solution.

Experts will be asked to rate tools and contents by means of subjective metrics. A set of educational tools and material from the MIREIA platform with regard to each learning purpose (medical anatomy, laparoscopy and/or flexible endoscopy) will be selected by the Consortium to be used in this validation process.

A subjective questionnaire will be developed such that it will allow to rate aspects of the educational tools and materials using a 5-point Likert scale (Strongly Agree, Agree, Neutral, Disagree, and Strongly disagree). These experts who will participate in this validation will not have been involved during the development stage of the educational materials and tools to avoid biased evaluations.

Six examples of items to be evaluated by the experts include:

- The educational tools and contents are appropriate for the learning purpose (medical anatomy, laparoscopy and/or flexible endoscopy).
- The variety of educational contents is appropriate.
- The educational value of the materials is adequate.
- The usefulness of the tools is adequate.
- The tools have good technical quality.
- The educational materials are adequate for improving the knowledge in human anatomy, laparoscopy and/or flexible endoscopy.

3.2. Usability validation

Usability validation mainly assesses the attractiveness and user-friendliness of the solution (MIREIA Platform). A questionnaire will be developed about the design, the structure and navigation of the project solution. The **system usability scale (SUS)** will be employed as a valid methodology to measure subjective usability [Brooke1996]. It provides a 10-item questionnaire to users, which the user must rate on a Likert scale from 1 to 5 (Strongly agree, Agree, Neutral, Disagree and Strongly disagree). To avoid repeated scores, odd questions ask about positive aspects and even questions about negative ones.

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

SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own. To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1,3,5,7 and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU. SUS scores have a range of 0 to 100.

3.3. Functionality validation

Functionality validation will assess aspects related to navigability and functionalities of the MIREIA platform. In this case, experts will be asked to rate these aspects using a 5-point Likert scale (Strongly agree, Agree, Neutral, Disagree and Strongly disagree). Statements will be such as the ones listed below:

- The access to the platform (login) is easy.
- The tools provided by the platform are user-friendly.
- The materials provided by the platform are easy to use.
- The educational materials provided by the platform are easy to download.
- The platform allows for different types of educational resources.

The subjective validation will be conducted during the M23 (November 2022) and M26 (February 2022). A report with the results of this subjective validation will be provided by M26 (February 2022).

4. Objective validation

For the objective validation, a set of objective parameters will be assessed at CCMIJU, MEDIS, SINTEF/St. Olavs hospital, and OUS institutions of the developed tools and contents in MIREIA project, both virtual and physical 3D models. Medical students, residents, physicians, and/or surgeons will be invited to participate as users in this validation. Depending on the particular profile of the participant, he/she will take part in the validation of one or more case studies (Virtual tasks and/or Physical 3D models). In addition, user who participated in the subjective validation could also take part of this objective validation. In this sense, we can optimize the user participation.

Table 4. Minimum number of participants per institution and use case for the objective validation.

Use case / Partner	Virtual tasks	Physical 3D models
CCMIJU	30	30
MEDIS	30	30
St. Olavs	30	30

OUS	30	30

The different tools developed for MIS training will be validated. These tools are:

1. Virtual tasks
2. Physical 3D-printed models for MIS using a box trainer

The virtual tasks and 3D-printed models for MIS training will be defined previous to the execution of the objective evaluation by the involved partners. The virtual tasks will allow users to learn basic anatomical knowledge in medicine and minimally invasive surgery. In the case of 3D-printed models, they will allow for the training of basic medical/surgical technical skills.

External end-users (students, residents or experts) will be asked to make use of the educational material (virtual tasks or 3D-printed models). The user will be asked to perform the virtual task medical/surgical training. This will be a basic training task so that it could be accessible to any level of experience. Regarding the 3D-printed model, it will be designed to be used for diagnosis or hands-on medical/surgical training.

For each scenario defined for this objective validation, a set of possible objective metrics will be defined, such as execution time, accuracy in the diagnosis, number of interactions with the virtual model, number of performance errors, etc.

During the data analysis, metrics obtained for the metrics obtained for groups of different experience levels in MIS could be compared (construct validation). This will provide information about whether the analyzed metrics using the provided training tools are able to distinguish among levels of expertise.

The objective validation will be conducted during the M23 (November 2022) and M26 (February 2023). However, given that the development of tools and results in the framework of the project is an ongoing process, it is likely that the validation process will be extended throughout the year 2023 in order to obtain more and more reliable results from the project validation. A report with the results of this objective validation will be started at M26 (February 2023).

5. Data Analysis

A descriptive and statistical analysis of the data extracted from the subjective evaluation and the objective evaluation will be performed. Table 5 summarizes some of the measurements that will be taken into account for each of the dimensions considered in the different validations.

All tests will be conducted with statistical analysis tools such as R, SPSS or MATLAB®. A critical qualitative analysis will be performed to discuss the main advantages, limitations and future challenges of the MIREIA platform.

Table 5. Expected end users in the validation of the final release.

Dimension	Measurements
Usability validation	- Descriptive statistics - Significant differences between demographic groups (t-student/Mann-Whitney; ANOVA/Kruskal-Wallis, etc.) - SUS score [Brooke1996]
Functionality validation	
Content	

Objective validation	- Execution time - Number of errors - Number of interactions - Etc.

A report with the results of this data analysis will be provided by M36 (December 2022).

6. Timeline

	2022				2023											
	Sep	*Oct*	*Nov*	*Dec*	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
Technological verification		D7.2														
Subjective validation																
Objective validation						D7.3										
Data analysis																D7.4

D7.2 Report for the technological validation

D7.3 Report for the subjective and objective validation

D7.4 Report for the data analysis

7. References

Brooke J (1996) SUS: a 'quick and dirty' usability scale. In: Jordan, PW, Thomas, B, Weerdmeester, BA, and McClelland, IL (Eds.). Usability Evaluation in Industry (189-194). London: Taylor and Francis.

Colardyn D, Bjornavold J (2004) Validation of Formal, Non-Formal and Informal Learning: policy and practices in EU Member States. European Journal of Education 39 (1):69-89.

Thalheimer W (2016). Performance-focused smile sheets: A radical rethinking of a dangerous art form. Hillcrest Publishing Group.