Interactive Music Science Collaborative Activities

Team Teaching for STEAM Education

Deliverable 6.1

Pilot testing action plan

Date:	30/6/2017
Author(s):	Fani Stylianidou (EA)
Contributor(s):	Erica Andreotti (UCLL), Colette Laborde(CABRI), Vassilis Katsouros (ATHENA), Fotini Simistira (UNIFRI), Manolis Chaniotakis (EA), Petros Stergiopoulos (EA)
Quality Assuror(s):	Renaat Frans (UCLL), Colette Laborde (CABRI)
Dissemination level:	PU
Work package	WP6 – Pilot testing in real settings and evaluation of educational value
Version:	1.0
Keywords:	pilot testing, action plan, in-school implementation
Description:	A practical guide to every aspect of piloting activities in accordance with the pedagogical framework and the technical progress of the project.

Horizon 2020
European Union Funding
for Research & Innovation

H2020-ICT-22-2016 Technologies for Learning and Skills
iMuSciCA (Interactive Music Science Collaborative Activities)
Project No. 731861
Project Runtime: January 2017 – June 2019

Executive Summary

The deliverable D6.1 'Pilot testing action plan' provides a concrete time-plan for piloting activities in the context of the EU project iMuSciCA. This commences in October 2017 and lasts until the end of December 2018. The testing will be undertaken in each of three European countries (Belgium, France and Greece) participating in the iMuSciCA consortium.

More particularly, this document aims to serve as practical milestone guide to every aspect of piloting activities in accordance with the pedagogical framework and the technical progress of the project, that is:

- set the context of the testing to be carried out;
- identify the research foci to be addressed;
- outline the research methods and instruments to be used;
- indicate the plans for data collection, including the envisaged distribution of effort towards the number of teachers and students involved in piloting;
- indicate the plans for analysis and reporting;
- include the ethical measures to be taken.

In the sections below, the main purpose of the pilot testing is considered as two-fold:

a) to test the iMuSciCA learning environment proposed and suggest improvements along the development process;
b) to demonstrate how the proposed iMuSciCA learning environment supports mastery of core academic content on STEM subjects (Physics, Geometry, Mathematics and Technology) alongside the development of deeper learning skills through students' engagement in music activities.

Version Log			
Date	**Version No.**	**Author**	**Change**
10-05-2017	0.1	Fani Stylianidou (EA)	Initial content
10-06-2017	0.3	Vassilis KAtsouros (ATHENA), Fotini Simistira (UNIFRI)	Sensor content
12-06-2017 & 16-06-2017	0.4	Colette Laborde (CABRI)	Comments and contribution about French implementation plans
16-06-2017	0.5	Erica Andreotti (UCLL)	Comments and contribution about Belgium implementation plans and evaluation methodology
19-06-2017	0.6	Fani Stylianidou (EA)	Comments addressed
20-06-2017	0.7	Fani Stylianidou (EA) Colette Laborde (CABRI)	Version compiled and sent out for internal review REview of document
21-06-2017	0.8	Renaat Frans & Erica Andreotti (UCLL)	Suggestions and clarifications on pilot testing process. Review of document
22-06-2017 26-06-2017	0.9	Fani Stylianidou (EA) Colette Laborde (CABRI)	Suggestions and reviewers' comments addressed. Final version. Review. Comments added.
30-06-2017	1.0	Vassilis Katsouros (ATHENA))	Submission of deliverable to the EU

iMuSciCA is an H2020 project funded by the European Union.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Description
PU	Public Report
WP	Work Package
ATHENA	ATHENA RESEARCH AND INNOVATION CENTER IN INFORMATION COMMUNICATION & KNOWLEDGE TECHNOLOGIES
UCLL	UC LIMBURG
EA	ELLINOGERMANIKI AGOGI SCHOLI PANAGEA SAVVA AE
IRCAM	INSTITUT DE RECHERCHE ET DE COORDINATION ACOUSTIQUE MUSIQUE
LEOPOLY	3D FOR ALL SZAMITASTECHNIKAI FEJLESZTO KFT
CABRI	Cabrilog SAS
WIRIS	MATHS FOR MORE SL
UNIFRI	UNIVERSITE DE FRIBOURG

1. Pilot testing timeline

Appendix 1 of this document presents the iMuSciCA pilot testing timeline, as well as the interactions between the pilot testing and the presentation of the project's relevant pedagogic outcomes in terms of deliverables.

Pilot testing will include two main phases of activities led by members of the consortium in the three piloting countries. Phase A will focus on usability and pedagogical relevance of the iMuSciCA proposed learning environment, and on the educational value of the first version of lesson plans. Phase B will focus upon the extensive use of the technological and associated pedagogic frameworks in real classroom settings in order to fine tune the capabilities of the former, finalize the latter, as well as assess the impact of both on learning.

The three foci addressed by pilot testing can be summarized as follows:

- *Technical usability and acceptance*: learning environment works in different devices; main actors know about its functionalities and how to operate it.
- *Pedagogical fit and value*: learning environment is useful; brings the STEAM inquiry-based pedagogy into the classroom, increases opportunities for collaboration, co-creation and collective knowledge building.
- *Learning fit and value*: learning environment personalizes learning process; motivates to learn more; helps to achieve deeper learning competencies.

Taking into account the Summer school break Phase B is furthermore divided in two sub-phases, B1 and B2. In each sub-phase more advanced lesson plans and educational scenarios will be implemented while teachers will have the chance to design their own lesson plans as they will be more and more familiarized with the system. The selected lesson plans for implementation will not only be in line with the curricula in the participating countries (see deliverable D2.1), but they will also demonstrate the unique potential of the iMuSciCA resources to cover a wide range of educational needs (e.g. in relation to STEM subject matters, school settings, student age groups) across different countries. Originally in English, they will be translated in French, Dutch and Greek and customized accordingly in order to become accessible to all participating students.

Here are the two main phases and sub-phases of iMuSciCA pilot testing:

1. iMuSciCA Pilot Testing phase A (October – December 2017)

Implementation in the 7th and 8th grades of secondary school in Belgium (first two years of secondary school called 'Middenschool') and possibly 9th and 10th grades; in grades 8th and 9th (13 to 15 year-old students) of school in France; and mainly in grades 9th and 10th (14 to 16 year-old students) of secondary school in Greece.

Focus on:

- *Technical usability and acceptance by teachers and students*
- *Pedagogical fit and value by teachers*

2. iMuSciCA In-school Pilot Testing phase B (April 2018 – December 2018)

2a) iMuSciCA In-school Pilot Testing phase B1 (April 2018 – July 2018)

Implementation in the 7th and 8th grades of secondary school in Belgium (first two years of secondary school called 'Middenschool') and possibly 9th and 10th grades; in grades 8 and 9 (13 to 15 year-old students) of school in France; and mainly in grades 9th and 10th (14 to 16 year-old students) of secondary school in Greece and possibly in the 7th, 8th and 10th grades as well.

Focus on:

- *Technical usability and acceptance by teachers and students*
- *Pedagogical fit and value by teachers*
- *Learning fit and value by students*

2b) iMuSciCA In-school Pilot Testing phase B2 (September 2018 – December 2018)

Implementation in the 7th and 8th grades of secondary school in Belgium (first two years of secondary school called 'Middenschool'), possibly 9th, 10th, 11th and 12th grades as well'; in grades 8, 9 and 10 (13 to 16 year-old students) of school in France; and mainly in grades 9th and 10th (14 to 16 year-old students) of secondary school in Greece and possibly in the 7th, 8th and 10th grades as well.

Focus on:

- *Technical usability and acceptance by teachers and students*
- *Pedagogical fit and value by teachers*
- *Learning fit and value by students*

2. iMuSciCA Pilot Testing Phase A

2.1. Introduction

Pilot testing Phase A (October 2017- December 2017) will focus on:

- *Technical usability and acceptance of the iMuSciCA learning environment, as perceived by teachers and students*
- *Pedagogical fit and value of the iMuSciCA learning environment and educational materials, as perceived by teachers*

2.2. Usability testing of first version of iMuSciCA platform

The usability testing of the first version of the iMuSciCA learning environment aims to direct the user-centered development process of the iMuSciCA prototypes. The usability testing will involve both types of user groups, i.e. teachers and students and the test goals are related to the user experience and the examination of certain features of the activity environments (AEs) of the iMuSciCA workbench. In particular, the usability testing will evolve around specifically designed tasks of the following AEs:

- Virtual musical instrument design
- Pen enabled canvas for music co-creation and interaction
- Performance/Interaction with musical instrument using gestures
- Music activities based on mathematical equations and geometric curves
- Performance sample sequencer co-creation and piano roll co-creation

In addition, common tools such as the visualizations of music signal as well as the handwriting input methods for mathematical equations shall also be examined.

The criteria of 5Es[1] (Efficient, Effective, Engaging, Error tolerant, Easy to learn) will be used to shape the measurements of the usability study. The goal setting of the 5Es are:

- **Efficient:** Can users find the information they need to complete tasks without assistance?
- **Effective:** Can users perform a process within a predetermined time frame?
- **Engaging:** Do users rate their experience as satisfying or enjoyable? Do their comments (and body language) suggest that they are having a positive experience?
- **Error tolerant**: Do users experience errors? And when they experience errors, do they recover successfully?
- **Easy to learn:** Can users get started right away? Does their ability to do tasks improve as they become familiar with the system? Does the system architecture match their mental model for the way they expect the system to work?

The users will be required to carry out specific tasks within scenarios of use and certain monitoring mechanisms will be set up to observe the user during her/his interaction with the iMuSciCA workbench. The iMuSciCA partners will present each participant the tasks one-at-a-time and instruct them along the procedure. At the end the participant will be given a questionnaire with ratings

[1] Barnum, Carol M., Usability testing essentials : ready, set...test! Morgan Kaufmann of Elsevier Inc., ISBN 978-0-12-375092-1, 2011, USA.

related to the tasks and the above mentioned criteria. Each participant, either student or teacher, will use the iMuSciCA learning environment in sessions of an hour long. The usability testing will be conducted with individual students in a dedicated room allocated for the needs of the study.

Observations of participants yield rich qualitative feedback. Qualitative feedback is gathered by noticing what participants do while they are engaged with the platform. Nonverbal feedback, such as users' facial expressions, eye gaze, and emotional state reveals a great deal about users' experience and can be captured with the setup of appropriate sensorial equipment. The following sensors and user monitoring mechanisms will be set up:

- Eye tracking: Use screen-based eye trackers to record user's eye gaze trace on the screen.
- Facial expressions: User facial analysis of webcam recordings of the user face.
- Galvanic Skin Response (GSR): Measure changes in the conductivity of the skin for inferring the emotional arousal and stress of the user.
- Electroencephalography (EEG): Record the EEG data to infer metrics related to engagement/boredom, frustration, meditation and excitement.

The above mentioned sensors will be installed and the user recording data will be captured and analysed with the use of the iMotions platform (https://imotions.com/).
The following equipment is necessary to be installed and properly running in the lab/room of the pilot testing:

1. User monitor with integrated webcam and speakers
2. Eyetribe eye tracker
3. Shimmer GSR
4. Emotiv EEG
5. Laptop with high specs for running iMotions platform

In addition, for the iMuSciCA workbench activities the students will use the following hardware devices:

1. Leapmotion
2. Kinect 2.0
3. Actuators for smart instrument

Finally, a high-speed internet connection is required to access the iMuSciCA learning environment.
Before conducting the actual usability testing of the first version of the iMuSciCA workbench, we shall hold trial testings on a small sample of participants of the sensorial monitoring equipment. The purpose of these trails would be to:

(i) identify the appropriate set up conditions
(ii) fine tune the calibration procedures
(iii) examine the quality of the recorded data taking into account possible artifacts
(iv) finalize the appropriate sensors and set ups for the pilot testings (Phase A and B)

Qualitative feedback also comes from participants' responses to questionnaire- and open-ended questions or interview after the end of the testing session. Their comments can provide rich insights into their experience, both positive and negative. Therefore, apart from the user monitoring mechanisms with sensors, users will be asked explicitly questions related to their rating evaluation of the 5Es of the iMuSciCA learning environment. A sample questionnaire can be found below.

Sample post-test questionnaire

1. Overall, please rate how easy or difficult it was to use the iMuSciCA web platform.
_ 1 — very difficult
_ 2 — somewhat difficult
_ 3 — neither difficult nor easy
_ 4 — somewhat easy
_ 5 — very easy

2. How easy or difficult was it to figure out the correct menus, icons, and procedures?
_ 1 — very difficult
_ 2 — somewhat difficult
_ 3 — neither difficult nor easy
_ 4 — somewhat easy
_ 5 — very easy

3. How much time do you think you would need before you felt proficient using iMuSciCA learning environment to perform similar tasks you were asked?
_ 10 to 30 minutes
_ 1 to 2 hours
_ 3 hours or more
Explain your response:

4. Please rate your overall satisfaction with the iMuSciCA learning environment.
_ 1 — very dissatisfied
_ 2 — somewhat dissatisfied
_ 3 — neither dissatisfied nor satisfied
_ 4 — somewhat satisfied
_ 5 — very satisfied

The team conducting the usability test will comprise of persons with the following responsibilities:

1. **Moderator:** This person has direct interaction with the participant and will guide her/him in the process of the test.
2. **Logger:** This person will be responsible for the setting up the equipment and also use the recording software of iMotions platform.
3. **Observer:** This person observes the participant during the conduct of the test and takes notes on her/his reactions.

In addition, there may be personnel that will take up the task of briefing the participant about the study and also collect the post-session questionnaire and conduct the open-ended questions or interview after the end of the testing session.

Team members conducting the testing will use the following checklist with the “to-do” items to remind the sequence of steps they have to go through before, during, and after each test session.

Checklist

Before participant arrives

- □ The equipment has been set up and work properly
- □ The iMuSciCA web activity environments are loaded
- □ Run a trial run of the tasks involved in the study

Welcome- consent form - instructions

- □ Introduce the participant to the purpose of the study
- □ Guide the participant to sit at the desk
- □ Show participant the location of recording equipment (EEG, GSR, cameras, microphone) and explain the recording mechanisms
- □ Go over the consent form, and ask the participant if s/he is comfortable with being recorded
- □ Give time for the participant to ask for questions or express his/her concerns

Guide through the test

- □ Explain the tasks and the activity environments of iMuSciCA
- □ Guide the participant through the tasks using the iMuSciCA platform
- □ Observers take notes of the participants reactions and use of the iMuSciCA platform

After completion, post-test questionnaire

- □ Give post-test questionnaire
- □ Conduct a brief post-test interview
- □ Thank the participant for participating in the study

Given the above mentioned setup of the pilot testing the 5Es are going to be measured and inferred from the observations and answers to questionnaire with the following techniques:
Effective: Go through the results of each task and measure the frequency of how often users have completed the task accurately and completely. Look for specific problems and mistakes that are made by several users and information that is skipped.

Efficient: Measure how long each task takes to be completed. Examine the eyetracking data and look for places where the screen layout or navigation makes it difficult for the user.

Engaging: Examine the facial/eyetracking/GSR/EEG data and look for signs that the screens are confusing, or difficult to read. Look for places where the interface fails to draw the users into their tasks.

Error Tolerant: Create a test case in which technical difficulties are likely to happen/appear, and see how well users can recover from these.

Easy to Learn: Control how much instruction is given to the test participants, or ask users to try especially difficult, complex or rarely-used tasks. Look for places where the on-screen text/icons or workflow helps or confuses them.

2.3. Pedagogical testing of first version of iMuSciCA educational materials

Based on the initial pedagogical framework and iMuSciCA use cases described in deliverable D2.1 and the initial evaluation criteria defined in deliverable D2.2, teachers in the three partner countries will be introduced to the STEAM pedagogy and asked to give their feedback on the initial educational scenarios and lesson plans of iMuSciCA (proposed in D2.3).
The aims will be to:

- familiarize teachers with the aims of the STEAM pedagogy;
- assess the pedagogic value of the proposed learning environment; and
- customize the proposed educational material to teachers', their students' and curriculum

needs.

Teachers who will have the opportunity to experience the preliminary version of the iMuSciCA learning environment will also be expected to give their feedback upon the project potential and perspective of designing their own lesson plans for the project.

2.4. Within Phase A Analysis

The analysis of the data collected in this phase will be reported in deliverable D6.2 ‘Interim report on teachers' feedback and pilot testing in schools’ as well as feed directly in the preparation of:

- the intermediate prototype of iMuSciCA’s learning environment for use by March 2018, as it will suggest technical improvements to be made;
- deliverable D2.4 ‘Intermediate Pedagogical Framework and use cases for teachers and learners’ (by end of December 2017), as it will suggest small scale adjustments on the approach and better targeted use cases for teachers and learners.

3. iMuSciCA Pilot Testing Phase B

3.1. Preparation for iMuSciCA pilot testing Phase B (January-March 2018)

In-between the two pilot testing phases, and more particularly in preparation for the in-school pilot testing sub-Phase B1, the following activities will take place:

1. finalization of the development of the intermediate version of the prototype of iMuSciCA's learning environment (January-February 2018) based on results of the analysis of usability tests with students and teachers (reported in D6.2);
2. organization of training workshops/webinars for teachers (February-March 2018)
3. finalization of monitoring system for pilot testing implementation (by end of February 2018)
4. first optimization of educational scenarios and lesson plans (by end of March 2018)
5. finalization of evaluation metrics (by end of March 2018)

3.1.1. Finalization of iMuSciCA's learning environment first prototype (January-February 2018)

During this period the iMusciCA technical team will make all the necessary development, controls and tests to make sure that the first integrated prototype of the iMuSciCA learning environment is fully functioning and ready for use.

3.1.2. Meetings, Workshops/Webinars with Teachers (February-March 2018)

In preparation of the in-schools pilot testing the three piloting partners will organize face to face meetings, workshops and/or webinars for teachers with two aims:

1. to introduce new teachers to iMuSciCA's learning environment and educational materials and recruit them to participate in pilot testing Phase B - induction meetings/workshops/webinars;
2. to prepare the teachers, who participated either in pilot testing Phase A or/and in the induction workshops and agreed to take part in the pilot testing of iMuSciCA, for this implementation – implementation meetings/workshops/webinars.

Both kinds of meetings/workshops are estimated to be of 2 to 3-hour duration and will take place in all three countries either face-to-face or via the web (as webinars).

The induction meetings/workshops/webinars will mainly focus on presenting to the teachers the STEAM pedagogy and the iMuSciCA learning environment. In all countries one or more of these meetings with teachers are expected to take place also during or even before pilot testing Phase A, in an effort to spread the information about the on-coming iMuSciCA school implementation and thus recruit the maximum possible number of teachers as early as possible.

These meetings and workshops/webinars aim at supporting teachers in the planning of their interventions and helping them achieve deep learning outcomes for their students. Some of the topics that may be discussed/presented are how to:

- use the iMuSciCA learning environment in the regular curriculum;
- adopt/create/develop new iMuSciCA lesson plans and educational scenarios;

- assess students for deeper learning; how to exchange best practices with other teachers.

The iMuSciCA piloting partners will also discuss with teachers what is expected from them in the framework of pilot testing, for example to use a series of iMuSciCA lesson plans in their classrooms, to create their own lesson plans using the iMuSciCA learning environment, to collect particular student data, etc. In case that additional support is needed the iMuSciCA partners will organize more personalized meetings and/or webinars with the teams in specific schools.

The iMuSciCA team will also offer opportunities for community building between the participating teachers in the same country during the full duration of the pilot testing phases, and after these until the end of the project. In the iMuSciCA approach such a community building approach is considered as a professional development activity that can help teachers to become progressively content providers.

3.1.3 Optimization of first version of Education scenarios and lesson plans (by end of March 2018)

Based on the results of pilot Phase A, but also taking into account the feedback from teachers attending the workshops/webinars organized during this period (and described above), the first version of education scenarios and lesson plans (D2.3) will be reviewed and enriched. These will be reported in the deliverable 'Intermediate Educational scenarios and lesson plans' (D2.6) by the end of March 2018.

In addition, after consultation with the teachers to be involved in the implementation, a selection of lesson plans/scenarios will be translated in French, Dutch or Greek accordingly and be made available electronically.

3.1.4 Finalisation of evaluation metrics for deeper learning (by end of March 2018)

Based on the results of pilot Phase A, but also taking into account the feedback from teachers participating in the implementation workshops/webinars the evaluation instruments and metrics to be used to measure the impact of the interventions and deeper learning will be finalized. These will be reported in the deliverable 'Final Evaluation metrics for deeper learning with iMuSciCA' (D2.5) by the end of March 2018.

3.1.5. Implementation reporting

During this preparation for Phase B, a robust reporting system will be set in place, allowing the consortium to have the full picture of piloting testing in-schools in the three participating countries. This reporting system will allow for a continuous feed to and reflection by the pilot testing coordinator in each participating country and the pilot testing coordinator of the iMuSciCA consortium (and leader of the relevant WorkPackage) at all times of how the piloting goals are met.

3.2. iMuSciCA Pilot Testing Phase B (April-December 2018)

3.2.1. Introduction

Pilot testing Phase B (April- December 2018) will focus on:

- *Technical usability and acceptance of the iMuSciCA learning environment, as perceived by teachers and students*
- *Pedagogical fit and value, as perceived by teachers*
- *Learning fit and value, as perceived by students and teachers*

3.2.2. iMuSciCA Pilot Testing sub-Phase B1 (April-July 2018)

The pilot testing in sub-Phase B1 will take different forms as explained below.

3.2.2.1. In real classroom settings

Two kinds of implementation of the iMuSciCA learning environment are envisaged as taking place in real classroom settings.

a) Lighter implementation

In this kind of implementation teachers will be expected to pilot test at least one to two lesson plans (worth of 2 to 4 teaching hours in total) with the same class of students within their regular curriculum hours. Of course they may also choose to implement the lesson plans with more than one class of students. This kind of implementation is not expected to provide evidence of deeper learning. However, it will provide data towards the three foci specified above.

The teachers who will take part in this kind of implementation will be expected to have participated at least in the induction workshops/webinars and more desirably also in the implementation workshops/webinars.

b) Heavier or in-depth implementation

In this kind of implementation teachers will be expected to pilot test at least 4 lesson plans (worth of 8 teaching hours in total) with the same class of students either within their regular curriculum hours, or in extra-curriculum school clubs. This kind of implementation is expected to provide data towards the three foci specified above, including evidence of deeper learning. The iMuSciCA's advanced behavioural and biosensor monitoring techniques will be used in instances of heavier in-school implementation to capture selected students' behaviour.

The teachers who will take part in this kind of implementation will be expected to have participated in the implementation workshops/webinars. They will also be the likely candidates to produce their own lesson plans.

Both kinds of implementation will take place in the three participating partner countries, though different effort will be invested in each one of them by each country as considered necessary and appropriate (see Chapter 4).

3.2.2.2. Student Summer Camp (June 2018)

In order to achieve an even more immersive experience of students with the iMuSciCA learning environment, the EA partner with the support of ATHENA is also planning to organize a student summer camp in Greece of 1 to 2-week duration (worth of 20-40 teaching hours in total) in which students will be involved in an iMuSciCA project, i.e. more than one iMuSciCA scenario. This is expected to provide data mainly about the 'Learning fit and value' of the iMuSciCA learning environment, addressing more closely its aspirations for deeper learning student outcomes. The iMuSciCA's advanced behavioural and biosensor monitoring techniques will be used more extensively to collect certain psychophysiological measures of students' activity, which will complement the evaluation of learning outcomes of the proposed pedagogical framework.

The student summer camp will take place in June 2018 after the end of the school year in Greece and will make use of an updated version of the first prototype of iMuSciCA learning environment. This updated version of the first prototype will have incorporated some of the feedback gained in the instances of the preceding lighter and heavier in-school implementation.

3.2.2.3. Summer School/Workshops/Webinars for Teachers

During pilot testing sub-phase B1 the two kinds of teacher workshops/webinars, induction and implementation, described in section 3.1.2 will continue, if not intensify, in order to support heavier implementations in schools, but also recruit and prepare teachers for pilot testing sub-phase B2.

In addition, a 1-week teacher training summer school will be scheduled by EA during the month of July 2018 with the aim to train teachers to use the tools of the iMuSciCA learning environment, to develop new lesson plans and scenarios, and share them with the larger teacher communities. As with the student summer camp above, the updated version of iMuSciCA's first prototype will be used in the teacher training summer school.

3.2.3. Within Sub-Phase B1 Analysis

The experience gained from the pilot testing sub-Phase B1 will feed in the preparation of the final deliverables of the pedagogical framework and use cases for teachers and learners (D2.7), and the education scenarios and lesson plans (D2.8), to be concluded by end of June 2018. These deliverables on the other hand will contribute significantly in the pilot testing sub-phase B2.

3.2.4. Between Pilot Testing Sub-Phases B1 and B2

Between the two pilot testing sub-Phases B1 and B2, the feedback gathered during sub-Phase B1 concerning the two evaluation foci 'Technical usability and acceptance of the iMuSciCA learning environment, as perceived by teachers and students' and 'Pedagogical fit and value, as perceived by teachers' will be analysed and used for:

a) the preparation of deliverable D6.3 'Final report on teachers' feedback and pilot testing in schools' to be concluded by end of September 2018, i.e. first month of pilot testing sub-Phase B2; and
b) the preparation of a running version of the iMuSciCA's learning environment, which will be the pre-final prototype envisaged.

3.2.5. iMuSciCA Pilot Testing Sub-Phase B2

The pilot testing sub-Phase B2 will mainly focus on the 'Learning fit and value' of the iMuSciCA learning environment, as perceived by students and teachers, although it is expected to keep providing feedback towards the 'Technical usability and acceptance' focus.

To these purposes both kinds of implementation foreseen, i.e. lighter and heavier, will continue. It is envisaged that a number of teachers and schools who have participated in the pilot testing sub-phase B1, in either kind of implementation, will continue to participate in this sub-phase B2. In addition, attempts will be made so that a minimum number of students will maintain an engagement with the iMuSciCA learning environment across the two sub-Phases, so that they can be monitored for deeper learning outcomes over a longer period. The iMuSciCA's advanced behavioural and biosensor monitoring techniques will continue to be used in appropriate instances of implementation.

The workshops/webinars for teachers will continue to take place as needed, changing progressively focus from one on induction and implementation to a broader focus on teacher professional development, as pilot testing will be drawing to an end.

3.2.6. Within and After Sub-Phase B2 Analysis

Feedback collected in this sub-Phase as well will continue to feed in the final stages of technical development of the iMuSciCA learning environment. Population of the iMuSciCA platform will also continue to take place as teachers will be making use and changing existing lesson plans and scenarios, as well as creating new ones.

Moreover, all quantitative and qualitative data collected concerning students' learning across pilot testing Phase B (B1 and B2) and using all kinds of research instruments, including the advanced behavioural and biosensor monitoring ones will be analysed as a whole and in complementary ways, towards preparation of deliverable D6.4 'Quantitative and qualitative results of learners', to be submitted two months after the end of pilot testing, that is at the end of February 2019.

4. Data collection, analysis, reporting

4.1. Overview of research methodology

As described in the DoA of iMuSciCA, the substantial part of iMuSciCA pilot testing will take place in real classroom settings, instead of test bed and lab environments, and thus the feedback gained will come directly from the end users. Evaluation in real school settings involving stakeholders in real time holds important challenges stemming amongst other things from the different contexts in which it will take place (e.g. different countries, educational systems, phases of education, etc.). This prompted iMuSciCA to follow a mixed-methods research approach adopting overall the interpretive paradigm. The interpretive paradigm offers means to closely document lived experience, foregrounding a socio-cultural framing considering three aspects: **personal** (focus on teachers and students and what they are doing), **interpersonal** (focus on interactions between teachers and students, and with peers) and **contextual** (institutional factors, teacher beliefs, resources, physical arrangements).

More particularly, iMuSciCA will gather quantitative and qualitative data from several sources and use pre- post- comparisons, where appropriate, to understand the impact of the intervention. The emphasis will be on understanding what teachers and students can do and how the proposed solution enables this, treating students as active, constructive, thoughtful meaning makers rather than in a deficit mode. Moreover, the focus will be on both the processes and outcomes of learning.

This research methodology is in a way also dictated by a) the kind of experiences promoted by the iMuSciCA learning environment and b) the deeper learning outcomes iMuSciCA wishes to accomplish.

The former are underpinned by the STEAM Inquiry-Based Science Education pedagogy described in deliverable D2.1, an educational approach which induces children to participate actively in school science. Important student benefits include: increase of motivation and confidence to learn; acquisition of process skills and scientific habits of mind; development of conceptual understanding, or learning of science content. These benefits may all be palpably apparent to teachers, but are not readily revealed on standardized measurements designed to test student achievement. Similarly, a number of iMuSciCA's expected impacts, such as the increase of students' engagement and motivation towards STEAM subjects and STEAM learning are not readily revealed quantitatively in the short truncated time that the implementation will last. Consequently, rather than relying solely on students' responses to pre- and post- questionnaires, student progression will be tracked through multiple measures of student work, including students' actions, talk and written products, as these will be recorded and analysed by iMuSciCA researchers using classroom observations and focus groups discussions; by the advanced behavioural and biosensor monitoring sensors iMuSciCA will employ; by the iMuSciCA learning environment itself using its analytics; and finally by their actual teachers through their reflections on their students' progress in focus group discussions and interviews.

An overview of the research methods instruments to be used, as well as about the data collection, analysis and reporting processes are given in the following sections. More specific details will be given in deliverable D2.2 Initial Evaluation metrics for deeper learning.

4.2. Samples

4.2.1. Testing sites, student age groups and teachers

The iMuSciCA's plan concerning the *minimum* number of students and teachers that will comprise the target samples in each country per Phase and kind of pilot testing is shown in Table 4.1. This plan has taken into account the different person-months effort foreseen for each partner for the relevant workpackage in the project's DoA. In addition, the assumptions made are that:

- teachers can be of Science, Mathematics, Technology or Music specialism;
- a number of induction and implementation meetings, workshops/webinars will take place in each country as necessary;
- the table below contains only the number of target teachers for the longer duration summer school and workshops in which teachers will develop iMuSciCA lesson plans;
- the in-school implementation mentioned in the table will be of both kinds heavier and lighter;
- at least 10 students in each country will participate in heavier in-school implementation in each of the Pilot Testing Phases B1 and B2.

*Table 4.1: iMuSciCA pilot testing **minimum** target samples*

PILOT TESTING PHASE A				
	Total	**GREECE**	**BELGIUM**	**FRANCE**
Students	**30**	10	10	10
Teachers	**6**	2	2	2
PILOT TESTING PHASE B				
Pilot Testing Phase B1				
In-School Implementation				
	Total	**GREECE**	**BELGIUM**	**FRANCE**
Students	**130**	60	50	20
Teachers	**6**	2	2	2
Student Summer Camp				
Students	**25**	25	--	--
Summer School / Workshops for Teachers				
Teachers	**40**	20	20	--
Pilot Testing Phase B2				
	Total	**GREECE**	**BELGIUM**	**FRANCE**
Students	**170**	80	60	30
Teachers	**6**	2	2	2

More details about what are the possibilities for pilot testing in each of the three piloting partner countries are described below.

a. Implementation in Belgium

Implementation during pilot Phase A will focus mainly on lower secondary education (7th and 8th grade(s), first two years of secondary school called 'Middenschool'). In these schools we find sometimes the new integrated subject STEM. Some classes have also a science subject called "Scientific Work" (in Dutch: "Wetenschappelijk Werk") where inquiry is at focus. Most classes have one (or two) hours of music education as well. All these subjects can form a suitable environment for

testing out iMuSciCA at this lower secondary level.
During Phase B1 and B2 the implementations in the lower secondary will continue and even be intensified but schools higher up can join the iMuSciCA piloting too (9th, 10th, 11th and 12th grade). At this upper secondary level, the courses of physics have 'waves and sound' on the curriculum and this might be combined with some hours of the so-called subject 'integrated seminar work' that some classes have in their program too. These seminars are organised around inquiry, they are rather interdisciplinary and the students are supposed to lead more or less their own work. iMuSciCA provides the right pedagogy for that.
In order to invite schools, UCLL organises an iMuSciCA information day on 20th of june 2017. On this day the STEAM pedagogy and some first outlines of iMuSciCA will be presented both in the lighter as in the full version. Teachers of interested schools can talk to the Belgian iMuSciCa team and if wanted, they can become candidate pilot school provided that they can fulfill boundary conditions like: an appropriate team of teachers prepared to implement the interdisciplinary STEAM pedagogy, give permission to smooth data collection and the like. UCLL will organise some training during the fall of 2017 and the spring of 2018 in order to make teachers capable of implementing iMuSciCA in the classroom.

b. Implementation in France

First contacts with schools have been established in France and proved to be very successful. The principle of an interdisciplinary approach combining in particular science and music has raised a high interest. The project pedagogy underpinning students' iMuSciCA activities was also very appreciated. Taking into account the French programme of studies about sounds, recommendations have been made mainly to carry out the pilot experimentation in lower secondary school (named « collèges » in French). Therefore, it is envisaged to conduct Phases A and B1 in lower secondary school (grades 8 and 9) and to experiment in upper secondary school in Phase B2 (grade 10). Changes in the policy of French education due to the change of government in France are about to occur during summer 2017 and may affect the way of organizing the experimentations within schools.

c. Implementation in Greece

Greece will follow both kinds of implementation (lighter and heavier). The choice of the implementation kind will be decided by teachers after consultation with the Greek iMuSciCA team. The selection may depend on the progress the classes will have made with respect to the curriculum in the second half of the school year (pilot testing Phase B1), as well as on the ICT resources available in the school. Particular attention, however, will be paid to the deeper learning process in order to ensure the appropriate conditions for a smooth and effective data collection procedure. The iMuSciCA implementation characteristics are shaped by the following three factors: testing facilities, student age groups and teachers' profiles.

For the planned implementation period B (B1 and B2) one could say that the kind of implementation (lighter and heavier) envisaged also corresponds in a way to the level of education provided by the school. Specifically:

In primary schools (upper primary) implementation can be performed in a computer room and/or a music room with laptops and internet connection. However, these are not readily available in all primary schools. The target student age group is mostly 10-11 years old, i.e. at grade 5 in which students get introduced to the most basic characteristics of music and sound and their scientific interpretation. Teachers at this level have no specialist knowledge about Science or Music and will follow the case studies framework for Natural Sciences as mentioned in the D2.1 Pedagogical Framework for Primary Schools. On the other hand, there are specialist Music teachers in primary schools, who can provide support and assistance. The choice of the lighter implementation kind is considered more appropriate at this level.

In lower secondary schools, implementation can take place in the framework of the national curriculum, in a computer room and/or a music room with laptops and internet connection. The student target age group is mainly 14-15 years old, i.e. at grade 9 in which the curriculum of Physics and Mathematics provide opportunities for heavy implementation. On the other hand, the lighter kind of implementation is considered suitable for grades 7 and 8, given the restricted opportunities offered by the respective mathematics curricula. The science, mathematics and music curricula for lower secondary schools in Greece are described in the D2.1 Pedagogical Framework for Lower Secondary Schools. In most cases the lower secondary school environment provides both specialist teachers for science/physics, mathematics and music. In any implementation the collaboration between the science/physics/mathematics and music teachers will be required. At this level of education, specialist 'Music Schools' also exist and will be targeted to participate in iMuSciCA (Special mention is made of these schools in D2.1). In addition, Ellinogermaniki Agogi's lower secondary school has committed to participate.

In upper secondary schools iMuSciCA will be implemented in two different kinds of school: 1) in a general education school - either in the classroom with laptops or in a special computer room, which have internet connection; and/or 2) in a specialist 'Music School' - either in a computer room or in special music-teaching facilities, which have internet connection. The student target age group ranges from 15 to 17 years old. The mathematics curriculum for these ages allows the teacher to explore the connection between the relevant areas and music in more detail. Again schools at this level will need to encourage the collaboration of mathematics and music teachers in accordance with the curriculum foreseen for both disciplines (please consult D2.1, 3.2.3 for "Upper Secondary"). Especially in 'Music Schools' the cooperation of the two teachers, suggested by the pedagogical framework, can ideally flourish as students follow, in addition to the formal curriculum, a special music programme for all musical instruments, during their extended school timetable. Actually a variety of music teachers specialized in different music skills in each musical instrument can support every research activity and deeper learning aspects scheduled by the action plan of the iMuSciCA project. Heavier implementation will be encouraged at this level of education, mainly in the framework of student clubs. Two upper secondary schools, Ellinogermaniki Agogi and the Music School of Pallini have already committed to such an implementation.

4.2.2. Identification of focus students for deeper learning

Given the restricted timespan of the evaluation process and the need to understand and assess children's progression in depth we will invite each teacher of the heavy implementation classes, if they have numerous students, to identify a focus group of **6-10 students per class** and make some close observations of their science learning over the testing period. It is recommended that a couple of students be selected for each of the three categories: more, less and much less experienced or confident as scientists and investigators. Generally it is wise not to include students with severe special needs or statements. The teachers will be asked to create a mixed gender group and with a range of digital skills, reflecting the diversity of their school community. These students will also take part in interviews and/or focus group discussion, conducted towards the end of iMuSciCA Pilot Testing sub-Phases B1 and B2.

4.3. Overview of research methods and instruments

4.3.1. Introduction

The research methods to be used in iMuSciCA pilot testing are appropriate to the research methodology chosen and described in section 4.1 above. Six levels of data will be collected: questionnaires, tests, classroom observation, reflections, behavioural and biometrical data and user analytics.

4.3.2. Questionnaires

Questionnaires will collect quantitative data from students and teachers concerning issues mainly of *Technical usability and acceptance* and *Pedagogical fit and value* foci a) during pilot testing Phase A; b) at the end of lighter and heavier implementations during in-school pilot testing Phase B; and c) at the end of the student summer camp and teacher summer school/workshops as appropriate.

In addition, an attitude questionnaire will be given to students prior to a heavier implementation of the proposed solution and then at the end of iMuSciCA pilot testing sub-phases B1 and B2. This will aim to detect any changes in students' motivation towards STEAM and STEAM learning, due to the intervention, and will be examined in conjunction with the behavioural and biometrical data about student engagement.

4.3.3. Content tests

In addition to questionnaires at the end of the interventions, students will be administered tests on the STEAM knowledge involved in the lesson plans, pre- and post- each implementation with the aim to measure any changes in student subject knowledge (i.e. *Learning fit and value*) as a result of iMuSciCA's use.

4.3.4. Classroom observation

Researchers will observe the real classroom implementation practice in the heavier implementation sites. Observations will mainly focus on issues of *Technical usability and acceptance* and *Pedagogical fit and value*; they will be looking for interactions between students and teacher and/or between students and the learning environment. It is suggested that mapping, sequential digital images, observation field notes using a timeline and audio recordings are amongst the most appropriate instruments to be applied in classroom observation. User analytics about students' activity are expected to complement qualitative data collected with the other methods.

4.3.5 Reflections

To further complement observations made in settings a certain level of reflection is required from both the students and teachers in the classroom. It is suggested that the digital images and audio recordings would be an appropriate means of triggering such reflections, and these could be elicited by way of interview and/or focus group discussions. Reflections will mainly address the *Pedagogical fit and value* and *Learning fit and value* foci of the evaluation.

Conducting interviews and/or focus groups are identified as possible areas for concern and some suggestions are given. For example, it is suggested that questions should be pitched at an appropriate level to put the interviewee at ease, with the initial discussion being led by the interviewee. Opportunities to ascertain teacher intentions should be sought, but with specific questions to guide this. Similarly, in the student focus groups questions should be pitched at an appropriate level for the students if researchers are to fully benefit from the focus groups sessions.
In acknowledgement of possible concerns raised, clear protocols will be developed to ensure that interviews and focus groups are conducted appropriately (i.e. ethically) and in order to gather rich and varied data.

4.3.6 Student behavioural and biometrical data

Visual attention and affective behavior in a learning platform such as iMuSciCA can have both positive and negative effect on students' learning. By investigating student gaze patterns and eye movements when interacting with the iMuSciCA web platform activity environments we can capture precise information about how discoverable or attention-grabbing visual elements such as navigation structures, screen graphics, links, text, or multimedia content are to the participants. Furthermore, confusion, frustration, boredom, as well as other affective states are elicited in response to the students' interaction with the iMuSciCA platform and they are inextricably bound to learning by affecting students' perceptions of the iMuSciCA environment and changing how well they learn from it. Towards this end, in order to assess student behaviour we shall employ the following monitoring mechanisms:

Eye tracking is the recording of eye position (gaze point) and movement on a 2D screen or environment based on the optical tracking of corneal reflections. Eye tracking reflects visual attention as it objectively monitors where, when, and what respondents look at. Furthermore, eye tracking devices can record the dilation and constriction of the pupil, which has been found to correlate with emotional arousal and cognitive workload. Eye tracking therefore can be used to validate and complement GSR measurements.
Facial expression analysis is a non-intrusive method to assess both emotions (subtle movements in face muscles, mostly subconscious) and feelings (accompanied by clearly noticeable changes in facial expression). While facial expressions can measure the valence of an emotion/feeling, they can't measure the associated arousal.

Electrodermal activity (EDA), also referred to as **galvanic skin response (GSR)**, reflects the amount of sweat secretion from sweat glands in our skin. Increased sweating results in higher skin conductivity. When exposed to emotional stimulation, we "sweat emotionally" – particularly on forehead, hands, and feet. GSR measurements can be done with lightweight and mobile sensors, which makes data acquisition very easy.

Electroencephalography (EEG) is a neuroimaging technique that measures electrical activity on the scalp. EEG tells which parts of the brain are active during task performance or stimulus exposure. It analyzes brain dynamics of engagement (arousal), motivation, frustration, cognitive workload and other metrics associated with stimulus processing, action preparation, and execution. EEG tracks stimulus-related processes much faster compared to other biometrics sensors.

Appendix 2 presents the metrics for these techniques.

4.3.7. User analytics

User analytics recorded by the iMuSciCA system will provide information about the interaction of the students with the learning environment.

4.3.8. Research instruments

Research instruments that correspond to the above methods will attend to students' voices as well as those of adults. The following instruments are suggested to be used to collect the same data across the three countries spanning all testing sites:

1. map of space and setting of computers
2. synthesis of personal fieldnotes with timeline
3. audio recordings and photos (and/or video, if appropriate) as verifying references
4. teacher questionnaire
5. student questionnaire
6. teacher focus group (or individual interviews)
7. student interviews or focus group
8. user analytics
9. student test
10. behavioural and biometrical sensors (only in Belgium and Greece).

4.4. Data collection

Suggestions as to the stage at which the above research methods and instruments might be used as well as an indicative framework categorizing methods against pilot testing foci have been developed and are shown in Table 4.2.

Table 4.2: Methods and Instruments Mapped against Pilot Testing Focus

Pilot Phase	Pilot Testing Focus	Research Instrument / Data Sources	Description
A, B	*Technical usability and acceptance in schools*	Student questionnaire Teacher questionnaire Classroom observation	*Student / Teacher:* Appropriate questions concerning the system's usability and usefulness, adequacy and relevance to users' requirements (incl. available user support) will be included in the questionnaires. Suggestions for improvement will be collected. *Classroom Observation:* During classroom observation, notes will be made of any related implementation issues occurred.
A, B	*Pedagogical fit and value*	Teacher questionnaire Focus group with teachers Classroom observation User Analytics	*Teacher questionnaire:* Appropriate questions concerning the system's contribution to the work of STEAM teachers and to curriculum enrichment and delivery will be included in the questionnaires. *Focus Group with Teachers*: Questions to these effects will be included in the focus group discussion. *Classroom observation:* During classroom observation, notes will be made especially in relation to teacher curriculum delivery, teacher-student interactions and monitoring of learning process. *User Analytics* will be used to establish teachers' use of the system for monitoring of learning process and assessment support and communication with students (if appropriate).
B	*Learning fit and value*	Teacher and Student questionnaires Focus group with students Classroom observation User Analytics	*Teacher and Student questionnaires:* Appropriate questions concerning the system's contribution to student's personalized learning will be included in the questionnaires. *Interview or Focus Group with Students*: Questions to this effect will be included in the focus group discussion. *Student test:* Pre- and Post- intervention STEAM knowledge tests aim to measure any changes in student subject knowledge.

			Classroom observation: During classroom observation, notes will be made especially in relation to effective and efficient individualized learning support by the system to students, as well as frequency of teacher-student interactions. *User Analytics*: will be used to establish students' activities on the system in and outside the classroom as well as the system's effective learning support.
B	*Learning fit and value*	Student attitude questionnaire Focus group with students Student behavioural and biometrical monitoring (Focus group with teachers & Teacher questionnaire)	*Student attitude questionnaire:* A questionnaire about students' attitudes to STEAM and STEAM learning will be completed by students prior to the start of the in-school testing phases, and then appropriately amended at the end of pilot testing sub-Phases B1 and B2. *Interviews or Focus group with students:* will include discussion about their self-concept of and motivation towards STEAM learning. *Student behavioural and biometrical monitoring data:* will provide information on students' affective states, elicited in response to their interaction with the iMuSciCA platform Relevant questions will also be included in the *teacher questionnaire* and *focus group discussion* with teachers.
B	*Learning fit and value*	Focus Students' grades Classroom observation Focus group with students Teacher interviews User Analytics	*Focus Students'grades:* the average results in relevant STEAM disciplines will be collected by the teacher for the 6 focus students prior to the implementation of phase B. *Classroom observation:* Notes made will include focus on the learning (concepts, skills and motivation) of 6 focus students in each class. *Interviews or Focus Group of Students:* Using a collection of photos taken from classroom observations as well as examples from students' work in iMuSciCA as prompts, students will be asked to comment on their learning. *Teachers' interviews:* will be organised in connection to the focus group, where teachers will be asked to reflect on the progress (conceptual learning, skills and motivation) of the 6-10 pre-selected focus students.

			User Analytics: will be used to describe students' learning progress.

Moreover, the following guidance is suggested to ensure parity in data collection across testing sites and countries (Table 4.3).

Table 4.3: Indicative Research Process – iMuSciCA Pilot Testing Phase B: Heavier implementation

When	What the researcher does	Which research instruments
Before first visit	Send Student Attitude Questionnaire Ask teachers to identify focus students in each testing class	
First Visit	Collect focus students' average results in STEAM disciplines Map of space and computer availability Perform student pretest on student subject knowledge. Synthesis of fieldnotes guided by a template of i) observations of students activities (including informal occasional conversations with students) ii) comments of teachers. Audio recordings and still images are used as verifying references for the researcher. Informal conversations with teacher (after the lesson) Looking at focus students' work Collect behavioural and biometrical data	1: map 2: synthesis of notes based on personal fieldnotes 3: audio recordings and photos as verifying references 8: user analytics 9: student test 10: behavioural and biometrical sensors
Second Visit	Synthesis of fieldnotes guided by a template of i) observations of students activities (including informal occasional conversations with students) ii) comments of teachers. with Audio recordings and images used as verifying references for the researcher. Informal conversations with teacher (after the lesson) Looking at focus students' work	2: synthesis of personal fieldnotes 3: audio recordings and photos as verifying references 8: user analytics 10: behavioural and biometrical

	Collect behavioural and biometrical data	sensors
Third Visit	Synthesis of fieldnotes guided by a template of i) observations of students activities (including informal occasional conversations with students) ii) comments of teachers. with Audio recordings and images used as verifying references for the researcher. Give out teacher questionnaire Give out student test Give out student questionnaire Collect behavioural and biometrical data	2: synthesis of personal field notes 3: audio recordings and photos as verifying references 4: teacher questionnaire 5: student questionnaire 8: user analytics 9: student test 10: behavioural and biometrical sensors

4.5. Overview of data analysis and reporting

4.5.1. Introduction

It is anticipated that the fieldwork will yield a rich collection of data, however it is of utmost importance that the data are utilised in a way which allows the consortium to address the main pilot testing foci. It is acknowledged that in each piloting partner country there will be at least one heavier implementation site and at least 10 students monitored for deeper learning outcomes. Lighter implementation will take place in more sites and more students (see Table 5.1). Collected data will be presented in their own right as part of the country reports; as such it is essential that there is consistency in the way these are recorded and presented.

During each of the testing sub-Phases B1 or B2, there will be typically 3 visits to the schools participating in the heavier implementation (see table 4.3 '*Indicative Research Process – iMuSciCA Pilot Testing Phase B: Heavier implementation*'). We call a narrative episode the outcome of each of these visits and a case as constituted typically by the 3 narrative episodes (from the 3 visits according to table 4.3). Any observed narrative episode will be considered valid for the pilot provided that an evidence of minimum two types of data are available (e.g. questionnaires, observations, interviews, focus groups). Where possible, researchers should seek the views and thoughts of the students in addition to those of the teachers. Notes should be provided, which contain the key areas of interest specific to the focus of the episode.

In order to ensure dependability, an iterative process of responsive evaluation [Abma and Stake, 2001; Youker, 2005][2] will be used with aim of the threefold improvement of the (i) technical usability and acceptance (ii) pedagogical fit and value (iii) learning fit and value. Via responsive evaluation, attention is given to the reactions of the practitioners (teachers and students), in this case the results of the teachers focus groups, the observations of students activities (synthesis notes, see table 4.3), the results of the students and teachers questionnaires. The outcome of this responsive evaluation (with the teacher focus group) will be concrete suggestions to improve the technical usability, pedagogical fit and learning fit.

This process will run in phase B1 and will be repeated in phase B2 for the heavy implementation schools.

4.5.2. Interim analysis and reporting

An Interim report on teachers' feedback and pilot testing in schools (D6.2) will be prepared for the end of December 2017. This will include the analysis of data collected during iMuSciCA Pilot Testing Phase A, addressing mainly issues of *Technical usability and acceptance* and including some preliminary findings about issues of *Pedagogical fit and value.*

---

[2]Abma, T.A., & Stake, R.E. (2001). Stake's responsive evaluation: Core ideas and evolution. New directions for evaluation, 92: 7-22.
Youker, B.W. (2005). Ethnography and evaluation: Their relationship and three anthropological models of evaluation. Journal of Multidisciplinary Evaluation, 2(3): 113-142.

4.5.3. Final analysis and reporting

The final report on teachers' feedback and pilot testing in schools (D6.3) will be prepared for the end of September 2018 and will contain the analysis of the data gathered during sub-Phase B1 concerning the two pilot testing foci *Technical usability and acceptance* and *Pedagogical fit and value*. In particular, this analysis will be used to refine the interim report findings and strengthen the evidence concerning the *Pedagogical fit and value* of the proposed solution.

On the other hand, all quantitative and qualitative data collected concerning students' learning across pilot testing sub-Phase B1 and using all kinds of research instruments, including the advanced behavioural and biosensor monitoring ones will be analysed as a whole and in complementary ways, towards preparation of an internal report by the same date (i.e. end of September 2018). Finally, the official deliverable on quantitative and qualitative results of learners (D6.4) across the whole of pilot testing Phase B (B1 and B2) will be prepared for the end of February 2019.

5. Ethics

5.1. Introduction

As mentioned in D8.4 'NEC – Requirement No.4', with regard to ethical issues detailed and substantiated reference to the following requirements will be fulfilled:

a) to ensure the informed consent of research participants and/or students' parents or legal guardians, since the research is directed to underage students;

b) to ensure participants' anonymity research and protection of their sensitive personal data in accordance with the current legislation;

c) to introduce the option for respondents to discontinue participation at any stage of their participation;

d) to protect research participants from exposure to possible physical or psychological harm or discomfort or other adverse for them effects as part of their participation in the research process.

More particularly in relation to the requirement (a), since the research involves underage students, we shall ensure that all the research participants (parents, students and teachers) comprehend fully:

- the research study's objectives;
- the process in which they will be involved;
- the reasons why their participation is necessary;
- the optional character of their participation;
- the option to withdraw their participation at any point during the conduct of the research.

In relation to the requirement (b) to ensure the anonymity of the research participants and the protection of their sensitive personal data, we shall respect all the provisions under the current legislation (L. 2472/1997 about the Protection of the Individual and Processing of Personal Data), and describe fully the ways in which the anonymity of all involved in the research will be ensured in every stage of the research (collection, processing and dissemination) as well as the treatment of all research data as strictly confidential.

The following sections are also included in deliverable D8.4.

5.2. Ethical considerations

The iMuSciCA consortium members agree to carry out this research in accordance with Data Protection Regulations and will comply with Directive 95/46/EC to ensure correct handling of data and privacy. During the course of this project, consortium members will take all the necessary steps to ensure that all participants (teachers, students, parents or legal guardians) understand the objectives of this project and the processes employed during the two pilot testing cycles to achieve them.

All beneficiaries will follow local and national regulations regarding data protection and will obtain approval from local/national authority in charge of data protection if applicable. iMuSciCA consortium will collaborate with the relevant educational authorities in the participating countries to receive favourable opinions of the relevant ethics committees and, if applicable, the regulatory approvals of the competent local and school authorities in the four participating countries. Copies of the official approvals from the relevant national or local ethics committees will be provided to the EC prior to the start of the respective pilot phase.

Many lessons and activities may be photographed or video-captured for evaluation of the impact of the proposed activities in the classroom. The data will be used only for evaluation purposes. It has to be noted also that all schools have specific rules on ethical issues that the consortium fully recognises and respects. Images, videos and other relative data from the research will be used for dissemination purposes, only if special permission from all relevant stakeholders is granted for this.

Assurances will be given about the confidentiality and anonymity of the data, with the removal of names and any identifying information, to meet the concerns of the child and responsible adult. Relevant written information will be provided for the child and responsible adult, and a contact telephone number, should they wish to contact the researchers. Moreover, the extent of the anonymity and any potential areas where the confidentiality of the interview may be broken will be explained to the child, e.g. at the outset of the interview. The researcher has also a duty to take steps to protect the child or other children, if they are considered to be 'at risk of significant harm'. The child needs to know what action may be taken in the event that she/he discloses that they or others are at risk of "significant harm", or where the researcher observes or receives information of incidents likely to cause harm. Arrangements will be made in advance, following professional advice, on agreed procedures in these cases, and for support for the child.

Information will be given about the storage of data and who will have access to it, and how it will be used, in the same clear language as used about the research. It can be argued that use by secondary researchers is not greatly different from that of primary researchers on the research team, who have not been directly involved in the collection of the data.

The evaluation team will use an internal, password protected area where only specific members of the consortium will have access (this is a common practice in the research projects in which the members of the consortium are involved) for the storage of all evaluation data. This internal system is capable to track the activities of the registered users (e.g. access, use of the resource, downloads). For example, the downloading of specific files (e.g. data or images) may not be permitted and the registered users can access them only on the server. The iMuSciCA consortium will nominate a member of the team to act as Data Manager to ensure the safe storage of all evaluation data, particularly where details of students are concerned. These data will be treated following the rules described above.

This project is focussing to identify the impact of the proposed activities on increasing students' achievement and interest in science, mathematics and technology. So the evaluation team will be tracking the students learning paths during the inquiry process and (in some cases) video-taping the classroom in order to verify changes in the attitudes and in behaviours. Numerous students from different schools will be involved for these purposes. No specific selection of the teachers (neither of their students) will be done in the framework of the project. It would be also valuable to mention that the "low achievers" seem to receive great benefits from the implementation of such methods in the classroom. So this is a real challenge of the consortium, namely to demonstrate the significance of the proposed approach even in classes with numerous "low achievers".

Finally, it has to be noted that the consortium will report on the progress of compliance with the Ethics requirements under each reporting period.

5.3. Informed consent procedures

To ensure the informed consent of research participants, the iMuSciCA project investigators will draw up a form, which incorporates the comprehensive plan of research. Since the research involves underage students it is needed to ensure the informed consent of parents and/or guardians. The information sheet and the informed consent form of research participants are essential documents for the consultation procedure. A sample project information sheet can be found in Deliverable D8.4 Appendix I, while more generic project information, namely a draft flyer can be found in the Appendix of Deliverable D8.1-H - Requirement No. 1. A Parental Informed Consent template can be found in Deliverable D8.4 Appendix II and an Informed Consent Form for Teachers can be found in Deliverable D8.4 Appendix III.

5.4. Participation of children

Beyond the formal informed consent form required to be given by the parents and guardians of students, it is necessary to take care on the part of the researcher to harmonize the conduct of the Articles 3 and 12 of the Convention on Children's Rights, which provide that: a) in all actions concerning children, primarily the child's interests should be taken into account (Art. 3) and b) the right to children who are able to express their views, to express them freely, taking into account the age and maturity should be ensured (Art. 12).

Therefore, together with the letter of parental informed consent, additional provisions for the students will involve the following: (i) inform students about the purposes and research procedure, (ii) the anonymity and the confidentiality of data will be guaranteed, (iii) the participation is optional and on a voluntary basis and not obligatory neither affects their performance in school and (iv) the students are able to withdraw at any stage of the conduct of the survey.

5.5. Data protection

The protection of the privacy of participants is responsibility of all persons involved in the iMuSciCA project. Privacy means that each participant can control the access to personal information and is able to decide who has access to the collected data in the future. The participants will be asked to give their informed consent before any personal information is collected during the studies. It shall be ensured that all persons involved in the research surveys within the project understand and respect the requirement for confidentiality. The participants shall be informed about the confidentiality policy that is used in iMuSciCA.

iMuSciCA will adhere to anonymisation policy. Personal data will be saved on secured servers, and will not be available to anyone outside the Project's team, unless participants clearly state their agreement for sharing these data. In case that data of specific participants is used for illustration in scientific publications or demonstrations, additional consent of the respective individuals will be required. This consent can be revoked at any time and the right to revoke consent will be expressed to participants when entering the study.

5.6. Templates of informed consent forms

The document of the informed consent form of research participants should include the following essential information:

1. FEASIBILITY RESEARCH: Describe the purpose of the research project (2-3 paragraphs)
2. PROCEDURE OF RESEARCH: Explain the data collection techniques and recording instruments used and the time required by the participants to attend the survey (3-4 sentences)
3. EXPECTED BENEFITS FROM RESEARCH: Describe how one expects the research results will contribute to the promotion of education and science.
4. POSSIBLE RISKS / DIFFICULTY: Identify potential risks or difficulties that could arise during the research.
5. ANONYMITY / PRIVACY : Describe how the anonymity of the participants and the protection of their personal data is ensured.
6. REFUSAL / WITHDRAWAL: Refer to the right of the participants to refuse to participate in the research and / or to withdraw from it at any stage.

Following the above guidelines, templates of the iMuSciCA informed consent forsm for teachers and students participating in the piloting are presented in Deliverable D8.4 Appendices II and III respectively.

Appendix 1: Pilot Testing Time Plan

	2017						2018						2019					
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
MONTH (start of)	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Professional development activities for teachers																		
Pilot testing Phase A																		
Pilot testing Prep for Phase B (B1&B2)																		
Pilot testing sub-Phase B1																		
Pilot testing sub-Phase B2																		
Deliverables (end of)																		
D2.2 Initial Evaluation metrics for deeper learning	*																	
D2.3 Initial Education scenarios and lesson plans	*																	
D2.4 Intermediate Pedagogical Framework and use cases for teachers and learners				*														
D6.2 Interim report on teachers' feedback and pilot testing in schools				*														
D2.5 Final Evaluation metrics for deeper learning							*											
D2.6 Intermediate Education scenarios and lesson plans							*											
D2.7 Final Pedagogical Framework and use cases for teachers and learners										*								
D2.8 Final Education scenarios and lesson plans										*								
D6.3 Final report on teachers' feedback and pilot testing in schools													*					
D6.4 Quantitative and qualitative results of learners																		*

Appendix 2: Metrics for behavioural and biometrical iMuSciCA tools

1. Eye tracking metrics[3]

Eye tracking renders it possible to quantify visual attention as it objectively monitors where, when, and what people look at. Fixation and gaze points are the most prominent metrics in eye tracking literature. Gaze points constitute the basic unit of measure - one gaze point equals one raw sample captured by the eye tracker. For example, if the eye tracker takes measurements 60 times per second, then each gaze point represents 16.67 milliseconds. If a series of gaze points happens to be close in time and range, the resulting gaze cluster denotes a **fixation**, a period in which our eyes are locked toward a specific object. Typically, the fixation duration is 100-300 milliseconds. The eye movements between fixations are known as **saccades**. Take reading as an example, our eyes don't move smoothly across the line while reading but they jump and pause, thereby generating a vast number of saccades. The visual span refers to how much text we can cover between fixations - on average, saccadic movements span 7 to 9 characters along the line of text. Typically, saccades are measured in angle velocity.

Heat maps: Heat maps are static or dynamic aggregations of **gaze points** and **fixations** revealing the distribution of visual attention. Heat maps serve as an excellent method to visualize which elements of the stimulus were able to draw attention -while red areas suggest a high number of gaze points and therefore an increased level of interest, yellow and green areas point toward flattening visual attention.

Areas of Interest (AOI): Areas of Interest (AOI) are user-defined subregions of a displayed stimulus. Extracting metrics for separate AOIs might come in handy when evaluating the performance of two or more specific areas in the same website or program interface.

Fixation sequence: Based on fixation position (where) and timing information (when) one can generate a **fixation sequence**. Dependent on where respondents look and how much time they spend, we can build an order of attention telling us where respondents looked first, second, third etc. Fixation sequence reflects salient elements (elements that stand out in terms of brightness, hue, saturation etc.) in the environment that are likely to catch attention. AOIs which respondents look at first are typically visually more appealing (more salient) and are therefore of more interest.

Time to First Fixation (TTFF): The time to first fixation is a basic yet very valuable metric in eye tracking and indicates the amount of time it takes a respondent to look at a specific AOI from stimulus onset. TTFF can indicate both bottom-up stimulus driven searches (a flashy label catching immediate attention) as well as top-down attention driven searches (respondents actively decide to focus on certain elements or areas on a website).

Respondent count: The respondent count allows to extract information about how many of the respondents actually guided their gaze towards a specific AOI. A higher count might indicate that fixations and gaze points are driven rather by external aspects in the stimulus material.

[3] Based on the Pocket Guide of iMotions (www.imotions.com) on Eye Tracking

Time spent: Time spent quantifies the amount of time that respondents have spent on an AOI. Since respondents have to blend out other stimuli in the visual periphery that could be equally interesting, time spent often indexes motivation and conscious attention (long prevalence at a certain region clearly point to a high level of interest, while shorter prevalence times indicate that other areas on screen or in the environment might be more catchy).

Pupil size/dilation: Pupil size primarily responds to changes in light (ambient light) or stimulus material (e.g. video stimulus). However, if the experiment can account for light, other attributes can be derived from changes in pupil size. Two common properties are emotional arousal and cognitive workload. An increase in pupil size is referred to as pupil dilation, a decrease in size is called pupil constriction. In most cases pupillary responses are used as a measure for emotional arousal. However, pupillary responses alone don't give any indication of whether arousal arises from a positive or negative stimulus.

Distance to the screen: Along with pupil size, eye trackers also measure the distance to the screen and the relative position of the respondent. Leaning forwards or backwards in front of a remote device is tracked directly and reflects approach-avoidance behavior. However, interpretation of the data is always very specific to the application.

Blinks: Eye tracking can also provide essential information on cognitive load by monitoring blinks. Cognitively very demanding tasks can be associated with delays in blinks, the so-called cognitive blink. However, other assertions can be derived from blinks. A very low frequency of blinks, for example, is usually associated with higher levels of concentration. A rather high frequency argues for drowsiness and lower levels of focus and concentration.

2. Facial expressions metrics[4]

Breakthrough research on computer vision and machine learning algorithms along with the gathering of high-quality databases of facial expressions all across the globe has been developed and made available to the broader public, technologies of instantaneously detecting faces, coding of facial expressions, and recognizing emotional states. These technologies use cameras embedded in laptops, tablets, and mobile phones or standalone webcams mounted to computer screens to capture videos of respondents face as they are exposed to content. More or less, all facial expression analysis engines comprise the following steps.

Face detection: The position of a face is found in a video frame or image, which can be achieved by applying the Viola Jones Cascaded Classifier algorithm. This might sound sophisticated, but one can actually find this technology in the camera of your iPhone or Android smartphone as well. The result is a face box framing the detected face.

Facial landmark detection and registration: Within the detected face, facial landmarks such as eyes and eye corners, brows, mouth corners, the nose tip etc. are detected. After this, an internal face model is adjusted in position, size, and scale in order to match the respondent's actual face. This is like an invisible virtual mesh that is put onto the face of the respondent. Whenever the respondent's face moves or changes expressions, the face model adapts and follows instantaneously. The face model is a simplified version of the respondent's actual face. It has much less details, (so-called

[4] Based on the Pocket Guide of iMotions (www.imotions.com) on Facial Expressions

features) compared to the actual face, however it contains exactly those face features to accomplish the task. Exemplary features are single landmark points (eyebrow corners, mouth corners, nose tip) as well as feature groups (the entire mouth, the entire arch of the eyebrows etc.), reflecting the entire “Gestalt” of an emotionally indicative face area.

Facial expression and emotion classification: Once the simplified face model is available, position and orientation information of all the key features is fed as input into classification algorithms which translate the features into Action Unit codes, emotional states, and other affective metrics. Feature classification is done for each emotion, Action Unit[5], and key feature independently, i.e. the classifier for smiles doesn’t know anything about the classifier for frowns, they are simply coded independent of each other. The good thing about this is that the automatic coding is accomplished much more objectively than manual coding where humans tend to interpret the activation of an Action Unit in concert with other Action Units, which significantly alters the results.

Automatic facial expression analysis returns numeric scores for facial expressions, Action Units, and emotions along with the degree of confidence. Most generally, you can think of these metrics as detectors: As the facial expression or emotion occurs and/or intensifies, the confidence score rises from 0 (no expression) to 100 (expression fully present).

The numerical values representing the facial expression can be plotted and visualized for both static images and videos (frame by frame time-course of emotion scores).

Take a look at the image on the left. The facial expression is not classified as a single emotion -instead, it returns the three emotions joy (99%), surprise (10%), and disgust (2%). The other emotions returned values of 0%. How come? Obviously, the current facial expression is a mixture of the three basic emotions. According to the classification, the expression most likely represents a highly joyful face with a slight indication for surprise.

The aggregation of facial expression data across respondents can be accomplished based on the actual values, but also on the thresholded and binarized signal of each respondent. This can be realized following the steps:

- Break the continuous data of each respondent during the selected stimulus period into intervals with a certain window length (500 milliseconds, for example).
- Compute the median of the facial expression score within each window.
- If the median exceeds the threshold, assign a value of 1 („true“). Assign a value of 0 („false“) to intervals where the data doesn’t exceed the threshold. Instead of the actual facial expression scores, the binary values 0 or 1 are used.
- In each time interval, sum up the binary scores across all respondents. For example, in case if one has collected facial expression data from 10 respondents and all of them showed a facial expression above the threshold in a certain interval, the aggregated value for this interval is 10. If none of the respondents had a facial expression that exceeded the threshold, the aggregated value for this interval is 0.

The resulting curve is a visualization of the consistency of your respondents’ facial expressions triggered by the provided stimulus material. The aggregated result relates directly to the

[5] Experts carefully examine face videos and describe any occurrence of facial expressions as combinations of elementary components called Action Units (AUs).

physiological and psychological arousal of the respondents set. One can also extract the number of occurrences of a specific facial expression, both within a single respondent as well as across all respondents.

3. GSR metrics[6]

The GSR signal consists of two main components:

Skin Conductance Level (SCL): The tonic level, known as skin conductance level (SCL), slowly varies and changes slightly on a time scale of tens of seconds to minutes. The rising and declining SCL is constantly changing within an individual respondent, depending on their hydration, skin dryness, or autonomic regulation. The tonic level can also differ markedly across individuals. As a result, the actual tonic level on its own is not that informative.

Skin Conductance Response (SCR): The phasic response rides on top of the tonic changes and shows significantly faster alterations. Variations in the phasic component are visible as “GSR bursts” or “GSR peaks”. The phasic response is also labeled skin conductance response (SCR) as it is sensitive to specific emotionally arousing stimulus events (event-related SCRs, ER-SCRs). These bursts occur between 1-5 seconds after the onset of emotional stimuli. By contrast, non-specific skin conductance responses (NS-SCRs) happen spontaneously in the body at a rate of 1-3 per minute and are not a consequence of any eliciting stimulus.

The following metrics are used in the analysis of the GSR data streams.

Number of GSR peaks: This metric counts how many GSR peaks occurred during a recording condition. One can measure the number of GSR peaks for each respondent and compute an average number across respondents or different groups of respondents.

Average GSR peak amplitude and magnitude: If present the same stimulus several times, ovε can compute the average GSR amplitude. For this measure, null responses (where the respondent showed no GSR peak at all) are simply ignored. In order to also reflect these, some authors additionally report the GSR magnitude, which includes null responses.

Since the GSR peaks across your respondents might vary in latency and/or amplitude, the aggregation cannot be a simple averaging of all GSR waveforms across respondents. Instead, one has to accomplish GSR aggregation based on binarization of the signal as follows:

- Break the continuous data of each respondent during the selected stimulus period into intervals with a certain window length (5 seconds, for example) and window overlap (1 second. for example).
- Assign a value of 1 (“true“) to intervals that contain at least one GSR peak. Assign a value of 0 (“false“) to intervals that contain no GSR peak. Instead of the actual GSR peak amplitudes, the binary values 0 or 1 are used.
- In each time interval, sum up the binary scores for a bin across all respondents. For example, in case if one has collected from 10 respondents and all of them showed GSR peaks in a certain interval, the aggregated value for this interval is 10. If none of the respondents had a GSR peak in the interval, the aggregated value for this interval is 0.

[6] Based on the Pocket Guide of iMotions (www.imotions.com) on GRR

The resulting curve is a visualization of the consistency of the respondents' autonomic arousal elicited by the provided stimulus material. The aggregated result relates directly to the emotional arousal of the respondents set.

4. EEG metrics[7]

EEG systems use electrodes attached to the scalp to pick up electric potentials generated by the brain. EEG electrodes are metal disks or pellets that connect with the skin via conductive gel, paste or cream, typically based on saline. The right combination of electrode metal and conductive paste is important as some metals corrode rather fast, resulting in poor data.

Under optimal conditions, the electrode and the electrode gel function as capacitor and attenuate the transmission of low frequencies (slow voltage changes in the delta frequency range, for example). The most common wet electrode type is made of silver (Ag) with a thin layer of silver chloride (AgCl) – you will often find descriptions like Ag/AgCl electrodes.

Once the voltage has been picked up by the electrodes, the continuous, analog signal has to be amplified and digitized in order to be stored on a computer. Although all of this happens under the hood, it's good to know some basic facts about amplification and digitization.

Since humans' brain is constantly active, there are continuous fluctuations and variations of the generated voltages. EEG systems, however, take discrete snapshots of this continuous process, generating samples of data - similar to pictures taken by a camera. EEG systems differ in the sampling rate (the number of samples per second) they can take.

If EEG is sampled at 500 Hz, samples are 1000 / 500 = 2 ms apart. High sampling rates can cause data overload as they require more disk space and longer processing and analysis times. Particularly if one is interested in frequency-based analyses (such as prefrontal lateralization of alpha or beta bands), a sampling rate of 128 Hz is more than sufficient.

Additionally to the digitization, the EEG signal is amplified. After the signals have been digitized and amplified, they are transmitted to the recording computer. This is either achieved through a wired connection (via USB, for example) or via Bluetooth.

As the electrodes will pick up electrical activity from other sources in the environment, it is important to avoid, minimize or at least control these kind of artifacts as best as possible. There are two types of artifacts: (i) the physiological artifacts that are related to muscle conductivity, eye movements, and blinks and (ii) the external sources of artifacts that are related to movements of the electrodes or headset movements, line noise, and swaying and swinging.

EEG analysis and feature extraction, involves a huge list of pre-processing steps in order to get results from raw signals. In fact, designing smart EEG paradigms is an art – analyzing EEG data is a skill. It certainly requires a certain level of expertise and experience, particularly when it comes to signal processing, artifact detection, attenuation or feature extraction. Any of these steps require informed decisions on how to best emphasize the desired EEG processes or metrics of interest.

What is a valid signal to one might be noise to anyone else. There simply isn't a generic data processing pipeline that one could apply to any EEG dataset, irrespective of the characteristics of the

[7] Based on the Pocket Guide of iMotions (www.imotions.com) on EEG

device, the respondent population, the recording conditions, the stimuli or the overall experimental paradigm.

Modern EEG systems come with an autopilot for data processing –they take the lead and apply automated denoising procedures or automatically generate high-level cognitive-affective metrics which can be used in order to get to conclusions much faster.

The ability to continuously monitor respondents' levels of fatigue, attention, task engagement and mental workload in operational environments is particularly useful in scenarios where wrong behavior could potentially result in hazardous situations. This information can be used to optimize software interfaces or entire environments that increase engagement, motivation and productivity.

Besides, the continuous extraction of psycho-physiological markers of engagement and vigilance from ongoing brain activity allows the design of closed-loop systems, which provide feedback on cognitive, affective and attentional states. In other words, whenever brain-based workload or drowsiness levels exceed a specified threshold value (or engagement levels fall below a certain value), respondents can be notified to initiate counteraction.

Emotive EPOC and EPOC+, that will be used in the study, belong to a family of EEG headsets that deliver cognitive-affective metrics. These devices allow you to collect the following metrics:

- Engagement/Boredom reflects long-term alertness and the conscious direction of attention towards task-relevant stimuli.
- Excitement (Arousal) reflects the instantaneous physiological arousal towards stimuli associated with positive valence.
- Stress (Frustration)
- Meditation (Relaxation)