ENVISAGE

ENhance **VI**rtual learning **S**paces using **A**pplied **G**aming in **E**ducation

H2020-ICT-24-2016

D5.2

Implementation of the educational scenarios and evaluation report

Dissemination level:	Public (PU)
Contractual date of delivery:	Month 12, September 30th, 2017
Actual date of delivery:	Month 13, October 24th, 2017
Workpackage:	WP5 - Pilot execution and evaluation
Task:	T5.1 Building virtual labs: Small-scale pilots addressing the educational scenarios, T5.2 System test and evaluation
Type:	Report
Approval Status:	Final
Version:	v3.0
Number of pages:	63
Filename:	D5.2 Implementation of the educational scenarios and evaluation report_Final version_v3

Abstract

The aim of this document is to present the implementation of the educational scenarios and an evaluation report for the delivered components within ENVISAGE. The deliverable reports on the results obtained during the execution of the implemented educational scenarios. The evaluation process focused on the three separate elements of the project 1) the authoring tool for building virtual labs 2) the analytics and visualizations tool for supporting the process of improving virtual labs and 3) the developed virtual labs as a means for successfully improving the learning process for teachers and students. The same three elements will also be subject for the second iteration required by the agile framework of the work package in month 21 for D5.4 (second phase).

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Co-funded by the European Union

Copyright

© Copyright 2016 ENVISAGE Consortium consisting of:

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the ENVISANGE Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

History

Version	Date	Reason	Revised by
V1	12-09-2017	First version	Line Ebdrup Thomsen og Benedikte Mikkelsen
V2	23-10-2017	Second version	Line Ebdrup Thomsen og Benedikte Mikkelsen
V3	23-10-2017	Review	Christoffer Holmgård
Final	24-10-2017	Final review/editing before submission	Giannis Chantas

Author list

Organization	Name	Contact Information
AAU	Benedikte Mikkelsen	benedikte@hum.aau.dk
AAU	Line Ebdrup Thomsen	let@hum.aau.dk
AAU	Anders Drachen	andersdrachen@gmail.com
EA	Georgios Mavromanolakis	gmavroma@ea.gr

Executive Summary

The current document aims to present the applied educational scenarios and to report on the results of the evaluation of the delivered components within ENVISAGE. The emphasis of the evaluation process was placed on the three separate elements of the project 1) the authoring tool for building virtual labs 2) the analytics and visualizations tool for supporting the process of improving virtual labs and 3) the developed virtual labs as a means for successfully improving the learning process for teachers and students. The deliverable reports on the results obtained during the execution of a number of small-scale pilots and works towards addressing if the requirements identified in WP1 are met. Succeeding to address the educational scenarios and requirements will reinforce the quality of the learning process for both students and teachers.

This deliverable builds on the methodologies and scenarios presented and discussed in D5.1. Most of the pilot activities were conducted by EA but supported by guidance and test material (e.g. surveys) provided by AAU. Subsequently, the data were analysed by AAU for informing redesigns and adjustment. The authoring and analytics tool were evaluated by teachers at the school Ellinogermaniki Agogi (EA). Through user testing and focus groups the teachers helped assess if the tools would support the development and decision-making when designing virtual labs. The outcomes of the evaluation were potential changes for the design and functionalities of both the authoring and the analytics tool, as provided by the test participants.

Abbreviations and Acronyms

PD	Professional Development
KPI	Key point of interest
UI	User Interface

Table of Contents

List of Figures

List of Tables

1 Introduction

The overall concept of ENVISAGE is based on the process of improving virtual labs through employing the authoring tool to build and improve instances of the virtual lab templates. Templates are templates within the authoring tool for developing new labs. The development process of the project is incorporated into four work packages (WP1-WP4), which are iterated over twice through the activities in WP5. The cycles start with the identification of the requirements and educational scenarios for the virtual learning spaces (WP1). Then they continue by using shallow game analytics for the aggregation of learner data and visualization (WP2) and carries on with predicting future and behavioural modelling, providing the appropriate learning content based on deep game analytics technologies (WP3). The cycle goes on with informing the process of decision-making when designing virtual labs using the authoring tool (WP4) and finishes with evaluating the degree to which the requirements gathered in WP1 have been fulfilled by the development conducted in WP1-4 (WP5).

ENVISAGE is designed as having a focus on maximizing the benefit for schools or companies using its assets. The benefit of the assets will, most likely, be identified when evaluating the tools. Evaluation protocols, using both qualitative and quantitative measures, will be utilized in the context of use cases, for ensuring that ENVISAGE succeeds in maximizing the benefit for its users. Based on discussions by the partners of the consortium, applicable use cases were selected.

1.1 The objective of WP5

The objective of WP5 is to conduct a number of small-scale pilots towards addressing if the requirements identified in WP1 are met. If successful in addressing the requirements, the quality of the learning process will be improved, for both students and teachers, when applying the tools of the project. Throughout the pilots the assets of ENVISAGE (the authoring tool and analytics and visualizations tool) are evaluated using appropriate educational scenarios. The evaluation for this work package focuses on the effectiveness of the developed technologies for optimizing virtual lab design and functionality and evaluates their ability to benefit educational organizations using the finished solution. The evaluation process will target three separate conditions**:**

1) The process of using the ENVISAGE authoring tool as a means for building virtual labs;

2) The support offered by the analytics tools in the process of improving virtual labs; and

3) The delivered virtual labs and learning content with respect to their effectiveness to meet the goals and expectations of both teachers and students in the learning process [1]

1.2 Structure of D5.2

The document will start with a presentation of the methodological approaches applied for each of the pilots as an evaluation protocol. Then it moves on to the presentation of the results from each of the pilots of all three conditions: the authoring tool, the analytics and visualization tool and the virtual labs. Here we present: how and which participants were selected for pilots, how the pilots were conducted and subsequently analysed. Lastly the analysis of the results from the pilot is presented. The last chapter of the deliverable is a summary and conclusion of what the pilots, as a whole, have brought to the project described individually for each of the elements developed. Furthermore, it will treat what future evaluations could cover within the ENVISAGE context.

2 Methodology and protocol of evaluation

2.1 Authoring tool

Two methods are applied to evaluate the authoring tool: heuristic evaluation and user testing. In this section, we briefly describe the theoretical foundations of each approach.

2.1.1 Heuristic evaluation

Heuristics are broad design guidelines, which can either be used for creating a user-friendly design or evaluating an existing solution, in order to increase its usability. Moreover, heuristics are used as a rule of thumb for either making decisions for new designs or for pinpointing weak points when evaluating existing ones. When used for evaluations, the inspection helps identify issues in the UI (User Interface) and is often performed by a group of reviewers analysing the interface based on the heuristics principles. This is typically called an expert review. For our expert review of the ENVISAGE authoring tool, we used Jakob Nielsen's 10 heuristics originally developed for interaction design [2]. This method was used because it is both cheaper and faster to conduct than traditional user testing. A heuristic evaluation therefore could be conducted before a user test and would help identify issues that could be corrected even before the user test. The method also requires multiple reviewers in order to yield the most valid results. For the heuristic evaluation two experts were used to gain more perspectives on the authoring tool usability, as they might have different understandings of how to use the authoring tool and thus represent different user groups. Additionally, using multiple reviewers has been proved to locate more issues in the tested software [2]. Additionally, ENVISAGE also had appropriately experienced experts on the team who could conduct a heuristic evaluation of the authoring tool.

Nielsen ten heuristics are described as follows [2]:

1. ***"Visibility of system status:*** *The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.*
2. ***Match between system and the real world:*** *The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.*
3. ***User control and freedom:*** *Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.*
4. ***Consistency and standards:*** *Users should not have to wonder whether different words, situations, or actions mean the same thing.*
5. ***Error prevention:*** *Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.*
6. ***Recognition rather than recall:*** *Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from*

one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. ***Flexibility and efficiency of use:*** *Accelerators — unseen by the novice user — may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.*
8. ***Aesthetic and minimalist design:*** *Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.*
9. ***Help users recognize, diagnose, and recover from errors:*** *Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.*
10. ***Help and documentation:*** *Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large."*

The heuristic evaluation was based on a list of tasks and a page-by-page approach, where possible violations for each of the 10 heuristics were investigated. The tasks were performed one by one, ensuring that all pages were inspected and that the reviewer got a sense of the workflow of the authoring tool. Two experts conducted the review and subsequently compared their individual results in order to decide which issues had the biggest impact and thus had been noticed by both experts.

Tasks followed by the reviewers for the evaluation:

- Create a new game project
- Create 3D scene
- Create 3D asset
- Drag and drop 3D assets to a 3D scene
- Save scene
- Edit an existing Scene (2D and 3D)
- Delete game project

2.1.2 User testing

When conducting a standard think-aloud test, the user is asked to perform a task while articulating thoughts and emotions occurring during the test [2]. The method is robust and flexible and is by some considered the most valuable method for generating insights on usability engineering [2]. It allows the collection of any user misconceptions about the UI design and is therefore a good fit for the ENVISAGE project. The method also works well in connection with additional methods like e.g. questionnaires, which can be used pre- or post user test depending on what is collected. In case of ENVISAGE, a set of standard questionnaires were used to evaluate the usability of the authoring tool.

Due to time constraints, not all user tests could be conducted on site with a facilitator sitting next to the test participant. Scenarios [3] were therefore developed to enable some participants to conduct the test remotely. This way, users could conduct the test at their own pace without feeling pressured by the presence of a facilitator. On the other hand, they were unable to ask for additional help if they needed it and notes from the facilitator could not be collected either. The scenarios however, assisted in making the tests more similar and comparable, as when using a test script, because all users received the same information. Results from the both the onsite and conducted remotely can therefore still be compared as the data has been collected using the same scenarios.

For this pilot, the test participants were provided with a set of tasks referred to as scenarios. The scenarios guided them through the authoring tool's different functionalities and ended when the participant had compiled a new virtual lab. During the test, participants were also asked to either think aloud, when performed onsite, or write notes, when performed remotely. Three of the user tests took place at the school Ellinogermaniki Agogi with a facilitator present, who took notes during the test, and for the two remaining tests the participants tested the tool, on their own, remotely. After the participants finished the scenarios, a set of standard questionnaires, see Appendix 6.4 and 6.5 , were provided to evaluate the authoring tool's usability and usefulness.

2.2 Analytics and visualization tool

The analytics and visualization tools developed for ENVISAGE were investigated using two other user research methods: focus groups and questionnaires.

2.2.1 Focus group

The focus group as a method is a means for studying values, attitudes, product preferences etc. [4]. The method requires a small group discussing a particular topic and uses the principles of group dynamics for achieving valid answers from its participants. The focus group is therefore also a qualitative approach for gaining recommendations on a topic or in the case of ENVISAGE, the quality of the visualizations picked by the designers.

For the first pilot, a focus group was conducted with science teachers of primary education at the school Ellinogermaniki Agogi. The test had three participants who were shown the visualizations previously presented in D2.3 "Visualization strategies for course progress reports" [5]. Visualizations were shown one-by-one, and the participants then had to articulate what they could read from the visualization and how easy it was to do so. The participants could also discuss their opinion when disagreements arose, providing us with even more insights. The collected participant statements on the visualizations were analyzed by the means of meaning condensation to enable only the most prominent findings to be summarized for future work.

2.2.2 Visualizations questionnaire

The method of questionnaires is often used for data collection, where big amounts of data need to be harvested from many users at once [6]. Additionally, the method is cheap as it only requires few resources to obtain a large dataset. However, the limitations of this method include the lack of direct communication with participants for gathering deeper insights and questionnaire fatigue [7]. The nature and number of questions were therefore tested internally before use, during the questionnaire's development.

For this pilot, the test participants were given a questionnaire with three metrics, each being visualized in two to three different ways. Participants were asked to rank the visualization internally with the metric and in relation to, e.g., best overview and most informative hereof. The participants were also encouraged to add more in-depth descriptions of why they had ranked the visualizations in the order they did. This pilot therefore helped us narrow down which visualization should be used for the final version of the analytics and visualization tool.

Ideally, the interface of the system should also have been tested during this pilot, however testing whether the visualizations were appropriate and meaningful for the end user was more important at this stage. In addition, at this point in the development and due to time constraints, the interface was not at a stage where it would make sense to user test it and hence it was decided to cover this aspect during the next iteration.

2.3 Virtual labs

The virtual labs of ENVISAGE were tested using a single user research method: a questionnaire.

2.3.1 Virtual lab questionnaire

For this pilot, the test participants were given five item Likert-like scale [8] questionnaire covering 18 questions evaluating the teacher's assessment of the quality of the lab. The participants were asked to consider statements that covered subjects such as the students' engagement with the lab, quality of educational contents, the fit in terms of the students' abilities and the teachers' expectations. The five-item scale ranged from "strongly agree" to "strongly disagree".

Ideally, the virtual labs could have been evaluated using sessions with teachers utilizing the labs in a class context where students interacted with them. However, at this point in the development and due to time constraints, this was infeasible, and the first iteration was evaluated by teachers using the virtual labs on their own. It could also be argued that if the teachers do not find the virtual labs quality to be of a high enough quality, the students would not either and this would be a waste of resources and students' time. For this pilot, the participants evaluated the lab based on their experience as teachers and completed the original questionnaire presented in D5.1 [9], adapted to suit the changed test situation.

3 Implementation of educational scenarios

This chapter is concerned with the presentation of how we implemented the educational scenarios by running a set of small-scale pilots. The pilots were designed towards evaluating each of the three conditions separately and the results will be utilized for an informed redesign of the elements in the delivered ENVISAGE services that could potentially be improved.

The conditions evaluated are: a) the ENVISAGE authoring tool as a means for building virtual labs; b) the analytics and visualizations tool, as a means for informing the process of improving virtual labs; c) the effectiveness of the learning content in the delivered virtual labs and their ability to meet the goals and expectations of both teachers and students in the learning process.

For the evaluation of the ENVISAGE authoring tool, two pilots were conducted. For the first pilot, usability experts performed a heuristic evaluation of the authoring tool in an effort to improve the user-friendliness of its user interface. For the second pilot, users evaluated their experience when engaged with the authoring tool, with the purpose of improving the user experience of the system. The results from the heuristic evaluation will subsequently be compared with the results from the user test in order to inform a redesign of the authoring tool and complete the development iteration. The two test methods provided us with a broader scope of insights, as users testing and expert review can yield different depths of inputs.

For the evaluation of the analytics and visualization tool, two iterations of pilots were also conducted. However, they used more similar methodologies than the two tests for evaluating the authoring tool. For the first pilot, three participants formed the basis for a focus group, conducted with the purpose of evaluating whether the selected visualizations were applicable for the end-users of the analytics and visualizations tool. However, some of the visualizations proved too complex for the users to read and additional new visualizations are thus needed. New visualization were identified, designed, and for the purpose of evaluating them a second pilot was conducted, realizing a quick second iterative step in response to the findings from the first pilot. For the second pilot, five users evaluated the new range of visualizations by ranking them compared to each other and based on different statements (e.g., most informative, best overview, etc.), in order to determine, which of the new visualizations could most efficiently support the teachers. The results from the two pilots were subsequently compared for this deliverable, in order to inform the second iteration of the visualizations tool.

For the evaluation of the effectiveness of the learning content in the delivered virtual labs, a pilot study was executed. The pilot consisted of a questionnaire distributed among teachers for evaluating to what extent virtual labs, developed using the authoring tool, would meet the goals and expectations of teachers. Subsequently, results from this pilot have motivated design changes in the virtual lab templates.

3.1 Heuristic test of authoring tool

The authoring tool has been evaluated through two different but complementary tests: a heuristic expert evaluation focused on usability and user tests with the intended end users. The following sections will describe the approach and the results of the heuristic test.

3.1.1 Methodological approach

For this pilot, usability experts performed a heuristic evaluation on the authoring tool in an effort to improve the user-friendliness of its user interface. A. C. Meyers Vænge, in Copenhagen, provided the context of the expert review and utilized the 10 usability heuristics originated by Jakob Nielsen [2]. The evaluation only focused on the finished functionalities of the authoring tool at the time, which were:

- Create a new game project
- Create 3D scene
- Create 3D asset
- Drag and drop 3D assets to a 3D scene
- Save scene
- Edit an existing Scene (2D and 3D)
- Delete game project

Using a linear task order use case, the heuristic inspectors went through the authoring tool page-by-page looking for violations of the heuristics. Inspections were conducted individually and consequently, notes are compared between inspectors to inform the report later in a later section. The results of the heuristic evaluation presented in this deliverable will only be a summary of the most prominent issues. The full report, as conveyed from the test team to the developers, can be accessed in Appendix 6.1 Usability Report: A heuristic evaluation of the authoring tool. Here all issues are reported and explained in detail.

The following functions have not yet been tested, as they were not yet in a state where such testing is applicable:

- Edit an existing scene (only 3D)
- Edit 3D asset
- Edit 2D scene (help, credits and main menu)
- Assemble and compile
- Delete game project
- Delete game scene
- Delete 3D asset

The functionalities listed above, which were not tested during the heuristic evaluation, were instead tested through the user test, as these were finished by then.

3.1.2 Test participants

For more information about the test participants, see Appendix 6.2 Participants for Heuristics test.

Table 1: Overview of test participant's demographics for heuristic evaluation.

Participant data	
Date:	Between 19-09-2017 and 21-09-2017
Where:	A. C. Meyers Vænge 15
Number of participants:	2
Age range:	26-38
Occupation:	Research assistants
Subjects:	Focus on usability and user experience testing and evaluation
Gender distribution:	Female, Female

3.1.3 Analysis of educational pilot

The most common usability issues in the authoring tool are violating the heuristics no. 10 (Help and documentation), no. 2 (Match between system and the real world) and no. 8(Aesthetic and minimalist design) as seen in the table below.

Table 2: distribution of usability issues across Nielsen's 10 heuristics in the authoring tool, as found by the inspectors

Heuristic:	Issue count
1. Visibility of system status	1
2. Match between system and the real world	7
3. User control and freedom	3
4. Consistency and standards	4
5. Error prevention	2
6. Recognition rather than recall	0
7. Flexibility and efficiency of use	1
8. Aesthetic and minimalist design	6
9. Help users recognize, diagnose, and recover from errors	0
10. Help and documentation	15

Help and documentation

Help and documentation is the most frequent issue reported during the heuristics evaluation. The heuristic covers issues related to providing users with the necessary information in an easy to search and user focused manner. Even though a system design that can be used without documentation is more desirable, help functions may still be required for users that are inexperienced.

In the case of the authoring tool, the issues reported were related to lack of help functions, tutorials, tool tips, user-guides, etc., informing the user how to utilize the tool. The lack of help is consistent throughout the entire authoring tool. However, this is to be expected for the first iteration of any system, as the core functionalities would be developed first (e.g. creating a new project, 3D scene or 3D asset) before help functions would be implemented. The evaluation also revealed missing previews, for functionalities such as Options, Help and Login when creating the main menu for the lab. Being unable to see a preview of the enabled functionality, the user will need to guess how the functionality works and what end user will see when playing the lab. In addition to missing previews for the functionalities in the main menu, the lab itself does not have a preview presenting the developer with the produced scenes. The developer would have to compile the whole lab, in order to view the scenes created and the 3D assets organized within. Compiling takes time, uses space on the computer and can be an exasperating process when e.g. fine-tuning the location of 3D assets. More previews, therefore, would ease the workflow of developers and document to them how their design will look for the finished version.

Match between system and the real world

Match between system and real world is the second most frequent issue reported during the heuristics evaluation. The heuristic covers issues related to the system conveying its content using the user language, words and phrases in order to familiarize the end user. Instead of using system oriented terms, the systems should follow real-world conventions and present information in a meaningful and logical order.

In the case of the authoring tool, the evaluation revealed ambiguous terminology, which could be misinterpreted and lead to a need for more help, as e.g. a tool tip explaining the function in the authoring tool. Instead of having tool tips, on all functionalities in the authoring tool, the solution was also to change the terminology to match the estimated end-users vocabulary better. The evaluation also revealed missing descriptions for elements in the authoring tool, leaving the users guessing how and when to use the functionalities. The user would then have to rely on previous experiences with similar systems, if the user has had any, and novel users will be highly challenged. The process of relying on previous experience requires a long and demanding thought process for the end user. Also, the goal of ENVISAGE is to accommodate the needs of teachers who are interested in the prospect of developing virtual lab supported by the authoring. Therefore, end users can be expected to be individuals with a relatively low level of experience using additional authoring tool or game engines.

Aesthetic and minimalist design

Aesthetic and minimalist design is the third most frequent issue reported during the heuristics evaluation. The heuristic covers issues related to dialogues not containing information, which the user would rarely need or is completely irrelevant. The information in dialogue should be considered carefully, as every unit diminishes the relative visibility of the other information units. Every extraneous unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

In the case of the authoring tool, the issues reported were related to unused top bars and links, irrelevant RSS feeds and minimum of characters. The unused elements will distract the user's attention from other elements that will probably be more relevant to the user. Another issue, with similar traits, is the banner of the front-page of the website. The picture used for the banner is a screen shot from a working version of the wind energy lab and on top a button saying, "learn more", an input field saying, "make your game" and two arrows pointing backwards and forwards. When a user locates an input field or button, they would expect being able to use them and the three elements therefore have a perceived affordance of interaction, as these type of elements would usually be interactive. The experience left with the user can be frustration and annoyance, which are emotions the ENVISAGE authoring tool is not attempting to generate in the users. Lastly, an RSS-feed is part of the webpage. While this might be relevant for some webpages, it is not the case for the authoring tool webpage. The RSS-feed would therefore be considered irrelevant and simply be one more unit of information fighting for the user's attention alongside more significant ones. In general, an overload of information and misleading buttons will steal the user's attention away from more meaningful elements, like e.g. starting the authoring tool or reading more about it. Avoiding the use of unexploited links, buttons or irrelevant information is hence recommended.

Least common issues

The least common issues are no. 6- Recognition rather than recall and no. 9(Help users recognize, diagnose, and recover from errors). Recognition rather than recall is related to minimizing the user's memory and not forcing them to remember information from one dialog to another. As the system architecture, is not vertically or horizontally deep, the user is not relying on remembering dialogs from one page to another. However, the heuristic also covers easy retrieval and visible instructions when suitable and the authoring tool has till now not implemented tooltips, instructions etc., as the core functionalities are more important to be implemented early. When instructions and tooltips have been implemented, more issues related to this heuristic is expected to arise. The second iteration is therefore expected to yield more Recognition rather than recall issues, which can then be reported and solve for the final version of the authoring tool. Help users recognize, diagnose, and recover from errors is related to error messages being explained in a plain language and not just, e.g., error 500, which novice users would not understand or know how to recover from. This kind of errors would probably be more apparent when the system is in a more finished state. The system is expected to come up with errors at its current state and might permit you to do something that you should not do later, which is perfectly fine when the system is under development. When the general workflows have been set up and the system is ready for the second iteration, it is also expected that more issues related to this heuristic will appear. The issues can therefore be addressed before the final version of the authoring tool is ready.

3.1.4 Summary

In short, the most common issues found during the analysis were related to missing previews, a lack of tool tips and help functions, a misleading terminology and missing descriptions of the authoring tool's functionalities. Issues of this nature are to be expected at this point of development and thus not a concern for reaching the goals of the project. Many of the issues are related to missing content or ambiguous terminology and will hence not require a substantial amount of recourses to correct. The developers of the authoring tool are therefore going through the issues presented in heuristics report and will then rename functionalities, changes the design and add more tool tips where it is needed, leveraging the findings from the heuristic evaluation to improve the next iteration of the authoring tool.

3.2 User test of authoring tool

This section will present the approach and results of the user test of the authoring tool.

3.2.1 Methodological approach

For the user tests the test participants were given a set of tasks in the form of scenarios that took them through the different functionalities of the authoring tool from creating a new lab, adding 3D objects to finally compiling a lab.

At the end of the user tests, the participants were asked to evaluate the authoring tool using a set of standard questionnaires focused on usability and usefulness, see appendices 6.4 and 6.5 .

Three of the user tests took place at the school Ellinogermaniki Agogi with a facilitator present, who took notes during the test. For the two remaining tests the participants tested the tool on their own remotely.

3.2.2 Test participants

The tests of the authoring tool were conducted with 5 test participants all of whom are teachers. The table below gives an overview the test participants.

Test Participant	Initial	Gender	Age	Teachers	Student age group	Test conducted
1	MC	Male	29	Physics, Astronomy	12-18	At EA
2	GMI	Male	42	Entrepreneurship	12-18	At EA
3	IA	Male	45	Physics, Sciences	10-12	Remotely
4	TT	Female	39	Physics, Sciences	10-12	Remotely
5	GM	Male	45	Physics, Math, Informatics	10-18	At EA

3.2.3 Analysis of Test Results

In this section, the results from the user tests are presented. As mentioned, the questionnaires that were used to collect the feedback from the test participants primarily focused on evaluating the usability and usefulness of the authoring tool. In the questionnaires, the teachers were asked to evaluate several statements based on Likert-like scale with response options ranging from unlikely to likely and strongly disagree to strongly agree as well as provide feedback in the form of comments.

Figures 1-3 give an overview of the given answers. With the respects to the limited number of test participants, the following will summarize the tendencies in the responses at an individual level rather than in aggregated statistics.

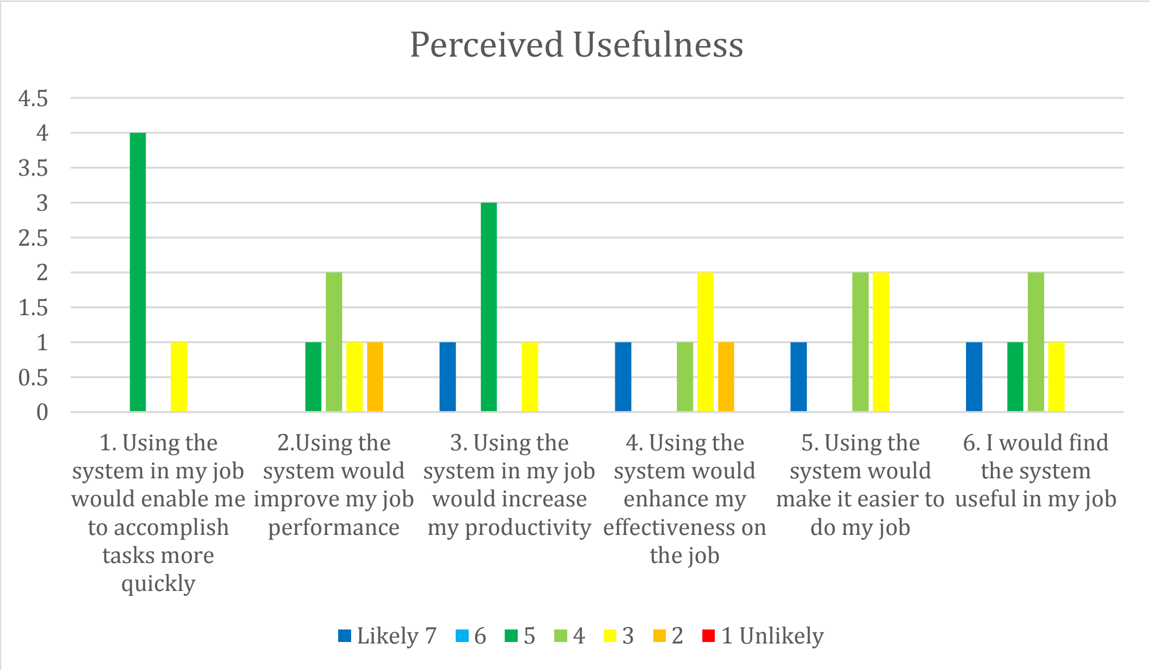


Figure 1 Summation of answers in relation to the perceived usefulness of the authoring tool.

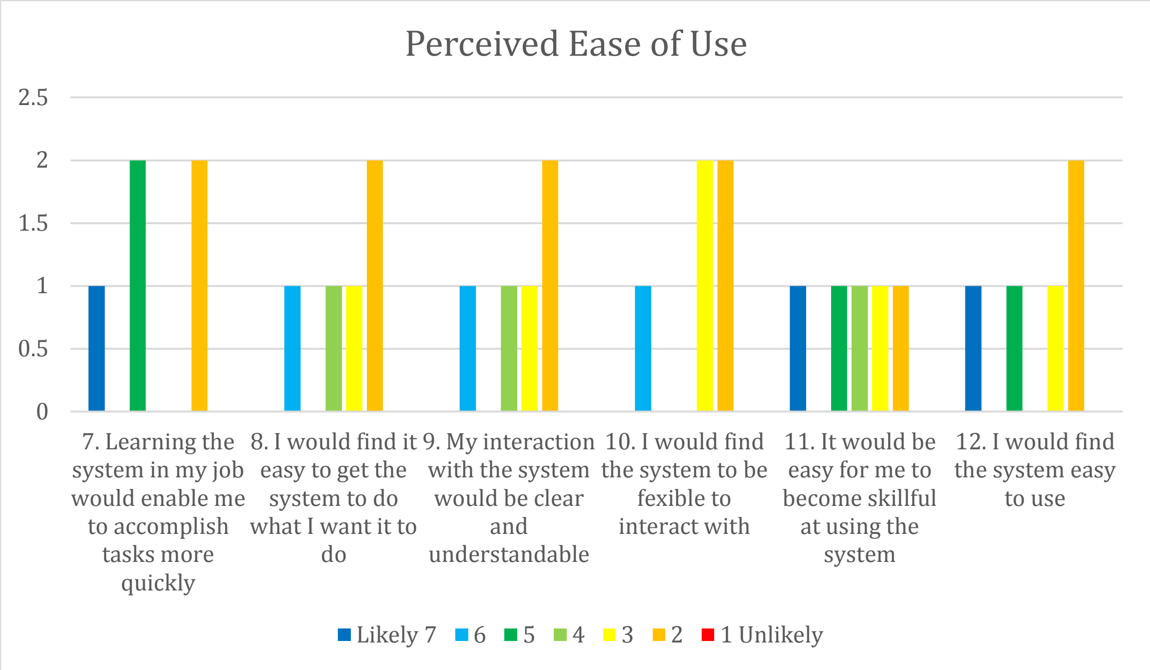


Figure 2 Summation of answers in relation to the perceived ease of use of the authoring tool.

Figure 3: The chart summarizes the testers' responses to the questionnaire on system usability of the authoring tool.

Comments on the test itself

From the comments that refer to the test material it is clear that although the testers found the task in the scenarios well described, they also found them too long and hard to complete and that testers redid tasks to get them right.

Usability

One of the participants consistently ranked the statements relating to the usability of the tool in a positive manner, while two had a more negative impression overall and the remaining two seemed more neutral in the response to the authoring tool's usability.

Three of the five participants made use of the comment sections. Their comments highlight different usability aspects both in negative and positive terms. The following will summarize these insights.

Interface

The questions that ask directly about the user friendliness of the authoring tool's interface are among the questions that get the most negative feedback. A sentiment that is also reflected in the comments with statements like "Not simple user interface" and "Arrangement of button" listed as the most negative aspects of the system.

Help functionality

Several of the comments described the need for extended help functionalities such as tool tips on the different elements of the interface, a user manual, tutorial or step-by-step guide to using the tool. Some of the testers commented on using test scenarios as the help function and noted that they were too long to be used as a form of help.

Navigating in 3D

Some comments specifically mention that it was difficult to interact with the 3D scenes. Rotating, scaling and placing objects as well as moving the avatar around in the 3D view is described as confusing and difficult. Some testers found the 3D view hard to understand and were unsure what the students would see in the compiled version of the created lab.

One of the testers was familiar with Unity, the game engine that the authoring tool is built upon, and made a comment that Unity is flexible and that the authoring tool should make better use of that flexibility.

Requested functionalities

A few functions were explicitly mentioned as missing from the authoring tool: The ability to undo actions, the ability to drag and drop objects into the 3D scenes and the option to make multiple copies of a whole lab.

Templates and pre-uploaded assets

Part of testers requested that more templates or examples should be added to the tool. Rather than creating the lab in a bottom-up manner, they could instead edit an existing template.

Along these lines are also the comments relating to uploading 3D assets. While one tester is happy to be able to add 3D assets found online to the lab others mention it would be helpful if instead of having to upload assets with three different files themselves to the authoring tool it were already stocked with assets ready to use.

Upload and Compile Time

The time it takes to upload assets and finally compiling the virtual lab was also commented on as taking too long. While it might be difficult to change how long these operations take, one could improve the user experience by adding additional feedback.

Usefulness

The questions that relate to the perceived usefulness of the authoring tool are harder to draw any direct conclusions from. In one case, it seems the tester instead of seeing it as an option to evaluate how the authoring could be used in their work as a teacher saw the questions as a potential critique of their work. For instance, the teacher responded to the statement "Using the system would enhance my effectiveness on the job" with the comment "I am already quite effective!".

However, there were also remarks related to the potential of the authoring tool and the experience of creating a virtual lab.

Potential

While there were quite a few comments that remarked negatively on the tools user-friendliness in the tested iteration there were also several comments that expressed a confidence that the authoring tool could offer benefits for the teachers going forward once more iterations are completed. One tester notes "I think it is a tool with a lot of potential", and another stresses the importance of having a simple tool that allows him to create educational content that can be adapted.

Sense of Accomplishment

Another positive comment that connects to the usefulness of the authoring tool is the sense of accomplishment the tool offered the testers. One teacher notes that without any prior experience with 3D editing he managed to create a realistic 3D game. Another writes it was a positive experience to be able to build a simple game.

3.2.4 Summary

As expected for a first iteration of user testing, the test participants encountered difficulties working with the authoring tool but at the same time they expressed an interest in the tool going forward and enjoyed the possibility being able to create a 3D experience for their classroom.

The most negative comments were with regards to the user-friendliness of the tool and the unfamiliarity of the 3D environment. Suggestions such as adding tooltips, more feedback from the system, and simpler a help functionality as well as an undo function could greatly improve some of the issues and should be considered in the further development

Adding pre-uploaded assets or sample template scenes for the users to edit and elaborate from rather than start from scratch would also give the user an improved experience and should also be considered for future versions of the authoring tool.

3.3 Analytics tool

3.3.1 Methodological approach

The focus group conducted with science teachers of primary education at the school Ellinogermaniki Agogi utilized the visualizations previously presented in D2.3 "Visualization strategies for course progress reports" [5]. The collected participant statements on the visualizations were analyzed by the means of meaning condensation to enable only the most prominent findings to be summarized for future work.

3.3.2 Test participants

For more information about the test participants, see Appendix 6.3 Participants for Visualization Test.

Table 3: Overview of test participant's demographics.

Participant data	
Date:	23-06-2017
Where:	Ellinogermaniki Agogi
Number of participants:	3
Age range:	39-45
Occupation:	Teachers
Subjects:	Physics, Sciences, Maths, Informatics
Gender distribution:	Female, male, and male

3.3.3 Analysis of educational pilot

The analysis of the educational scenario will now be presented. The structure of this section is organized as follows: a) a short description of the visualization b) a screenshot of the visualization, as seen by the participants c) the participant's responses for the visualization.

Visualization 1: Dashboard

A dashboard is used as an overview of KPI's connected and often also customized to fit a particular objective of its user. Linked to a database the dashboards can be updated constantly and is frequently used for websites to tack user retention, daily users, revenue, page views etc. The following dashboard is from deltaDNA.

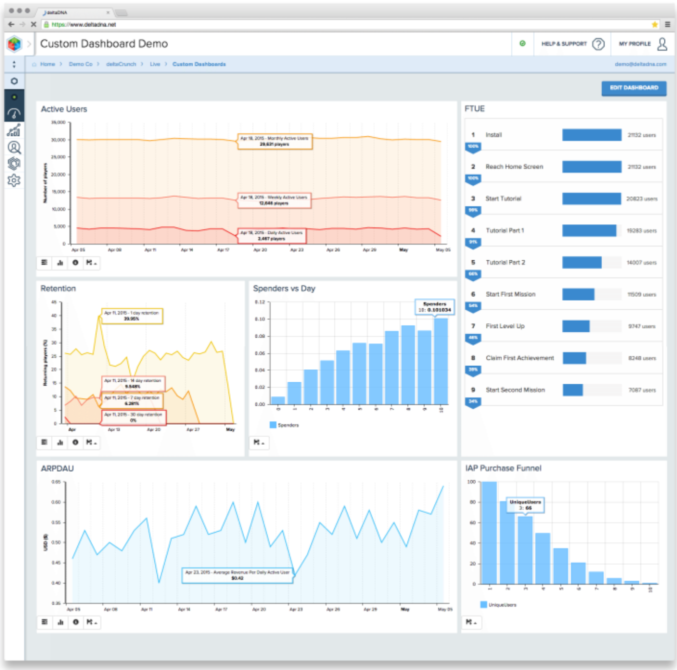


Figure 4: Visualization fromD2.3 [5], p.20, fig. 15 – example of a dashboard.

When shown the lab the teachers expected tabs or menus where they could select dates or classes displaying relevant information, save it and lastly print it. One teacher also said, “It is well received and is understandable intuitively”. The teachers also recommended that depending on screen size and resolution on the computer, maybe graphs should be arranged in sets of four or five where the user can scroll up or down to display more. This proposal was rooted in the teachers being aware of the rise in use of portable devices e.g. mobile phones or small tablets. At Ellinogermaniki Agogi, the students will soon be using tablets for their IT-supported classes and the teachers were therefore interested in being able to apply a potential dashboard for this device too.

Visualization 2: Bar chart

Bar charts use rectangular bars relative to the values they represent, for visualizing data with a categorical nature. The bars can be plotted both horizontal or vertically (like in Figure 5).

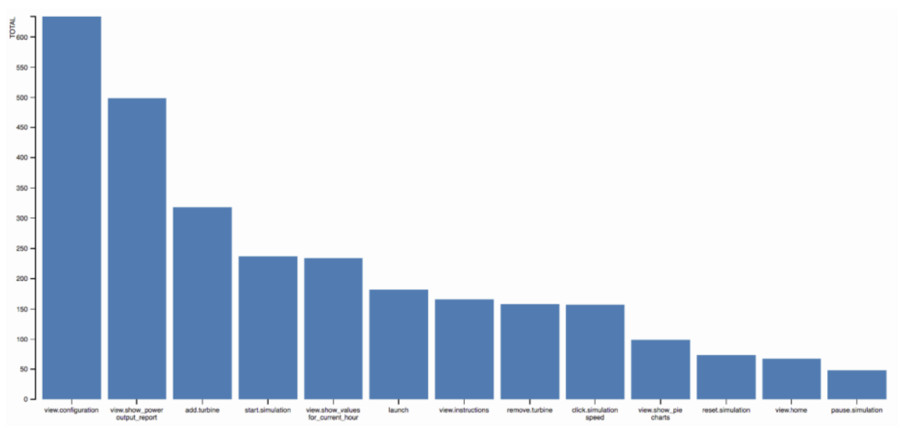


Figure 5: Visualization from D2.3 [5], p.24, fig.17 – example of simple bar chart.

The bar chart is one of the most basic visualizations for singular aggregated metrics within learning analytics. The teacher had the same perception of the visualization and deemed it easy to read but found the font size of the labels too small. One participant said, “It is very understandable intuitively. Size of text labels should be enlarged”. The visualization showed the number of times a unique event occurred in the dataset and subsequently ranks them relatively to each other’s value. The teachers quickly noticed the relationship between the X and the Y-axis, which could very likely be due to the rapid used of this type of visualization.

Visualization 3: Force-directed graph

In force-directed graphs, each node represents events, which can be tracked in the data. This type of graph is typically used for visualizing the travel paths between two data points like e.g. the sequence of two events. The weight of an edge (lines in Figure 6) between two nodes (dot in Figure 6) visualizes how frequently the sequence of the two events occurs in the data set used. The visualization is furthermore valuable for cross-sectional inspections of aggregated behavior in a group of individuals. This means that a teacher more easily could get an overview of how frequent a behavior is, or is not, among the whole class.

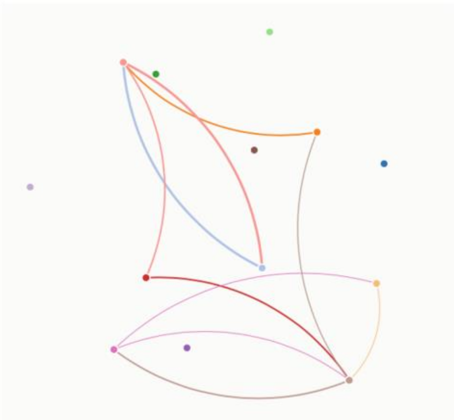


Figure 6: Visualization from D2.3 [5], p.25, fig.18 – example of a force-directed graph of a group of students.

The teachers understood the general concept of the graph as displaying a non-linear flow of actions performed by students but were not able to read anything directly from it. One teacher said, “Not intuitively understandable” and continued to request more guidance for how to read the graph “Extra information or guidance in textual or numerical mode may help to enhance comprehension and clarity”.

Visualization 4: Chord diagram

The chord diagram is a visualization method for demonstrating the transitions between frequencies between states in aggregate. The data is typically organized outwardly around a circle where the relationships between the data points are represented using arcs. The aggregation of data allows the inspection of e.g. the interaction patterns of a class.

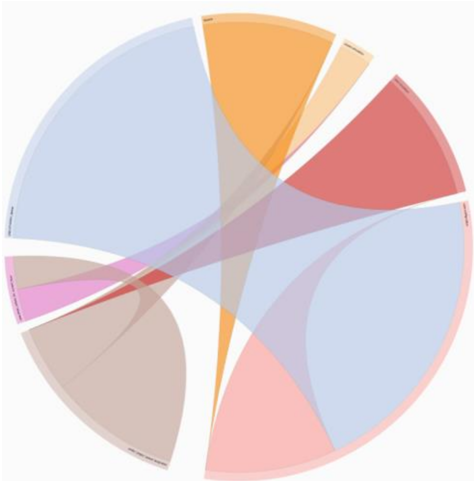


Figure 7: Visualization from D2.3 [5], p.26, fig.19 – example of a chord diagram.

Similar to force-directed diagram (Figure 6), the chord diagram is not intuitively understandable for the teachers, who are not sure how to read the graph, which also suffers from small text labels according to the teachers. Parallel with the bar chart (Figure 5), the chord diagram is also pointed out to have labels with a too small font size.

The teachers have been shown a paper version of the chord diagram, which subsequently have been made into a virtual and interactive version. Presented with this idea, one teacher felt that if she could interact with the graph virtually she would be allowed to comprehend the graph in detail then it would be much easier to interpret without difficulty.

Visualization 5: Absolute time-line

Timelines are usually used for visualizing a list of chronological events. Visually, the timeline is a long bar labelled with dates of the events. The dates are typically relatively scaled based on, e.g., time between the events. Absolute time-line displays the actual time spend on the task. The absolute time-line below leverages the full travel path metric and display which event each user performed, and how much time the users spend on each event. The timeline labels the time between events with the name of the event at the start of each period and displays information either for each user or for an average across groups.

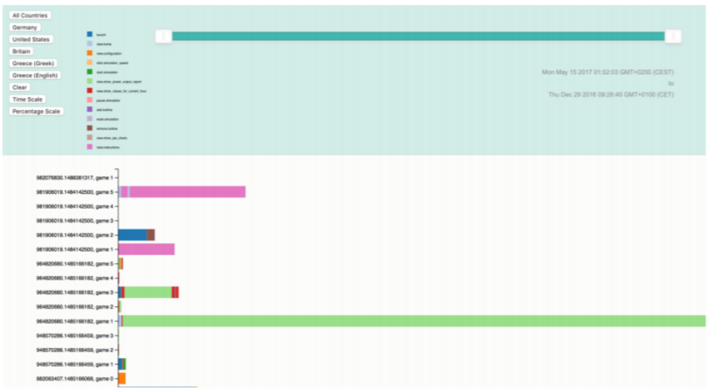


Figure 8: Visualization from D2.3 [5], p.27, fig.20 – example of absolute time-line view graph.

Teachers found the absolute time-line understandable intuitively but also noted that the text size of the labels for this visualization were too small. The teachers also requested more selection criteria's, such as, e.g., dates, classes, min/max time etc., which could be added in the menu. Again, the time-line is a rapidly used visualization, like the bar chart, which could be because the teachers intuitively know how to read the timeline.

The relative time-line visualizes an event sequence (e.g., travel paths) in the Wind Energy Lab for a group of students. Users travel paths are sorted by similarity, measuring the pairwise Levenshtein edit distances [5]. While the absolute time-line displays the actual time spend on the task, the relative time-line is time spent on a task relative to how much time other user spend. The relative time-line visualizes data sets with a high degree of variation better than the absolute time-line. If a student's time spent on a task has a high degree of variance it may become difficult to compare individuals or groups to one another, something which the relative time-line addresses more proficiently. In contrast, the absolute time-line is a more appropriate visualization strategy for data sets with a lower degree of variance.

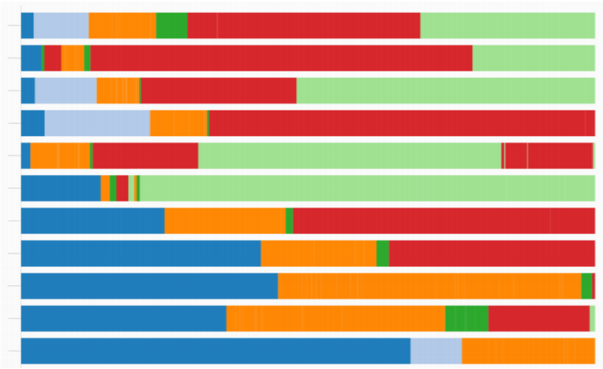


Figure 9: Visualization from D2.3 [5], p.28, fig.21 – example of relative time-line view graph. Color legend: Blue: time spent in launch; Light blue: time spent in home view; Orange: time spent in configuration view; Green: time spent in simulation view; Red: time spent looking at current hour values; Light green: time spent viewing power output report.

Similar to the absolute time-line, the relative time-line was also noted to be intuitively understandable but with too small text size of the labels for this visualization. For this time-line, the teachers also requested more selection criteria to be added in the menu.

3.3.4 Summary:

In general, the participants found four out of the six visualizations easy to understand, intuitively, but also articulated need for bigger text size of the labels, legends and menus. Because the visualizations were printed out and placed in front of the focus group participants during the discussion, some labels can have appeared smaller than intended by the designer. Another factor, which would probably also help address these issues, is the visualizations being converted into virtual and interactive versions. When the final visualization and analytics tool have been developed, the user would be able to interact with the visualizations more, thus allowing a deeper understanding of the content. Furthermore, most internet browsers allow scaling to support individual user needs for text size.

Additionally, the size of labels, legends etc. will be enlarged by the developers, when implementing the final visualizations.

The following pattern was also identified among the test participants: the teachers did not experience any issue reading the more commonly used visualization types (as bar charts and timelines) but encountered higher degrees of difficulty when reading the more seldom used visualizations. The pattern is not an unexpected result, as it is probably due to the strong mental models which the test participants have already built for reading the common visualizations, whereas, the chord diagram and force-directed graph are new concepts. This indicated that the visualizations used for the analytics and visualizations tool should be simpler than first expected.

3.4 Second Iteration Test of Analytics Visualisations

Based on the feedback from the first test of analytical visualisation it was decided to develop and test out a few different and simpler visualisations inspired by the data collected though the wind farm simulation.

3.4.1 Methodological approach

For this iteration, the test participants were sent three sets of dummy visualizations along with a short description of the wind farm simulation. The material was given as pdfs that could be printed so there was none of the interactivity the analytics front-end will offer. The intent was rather to get an idea of how the testers viewed different way of illustrating data without spending effort on implementation before the designs had been validated. Unfortunately for this additional test is was not possible to observe and interview test participants as they evaluated the visualizations all the feedback was given as written comments.

For each set of visualizations, the testers were asked to explain which example in the set they found easiest or hardest to understand, least or most useful, least or most informative, or gave the best of worst overview of the students' performance and, for two visualizations of the set, to rank the examples from best to worst.

The first set, titled Power Status, illustrate in three different ways the amount of time eight fictive students had the simulation either under, over or correctly powered. In the first of the examples, example A, the power status is shown in a horizontal bar chart. The second, example B, the power status is shown in individual pie charts and in the third, example C, the power status is depicted in absolute time in vertical bar charts. In an unfortunate hiccup the legend for example B was missing, however the colors were matched the legend in the other two examples.

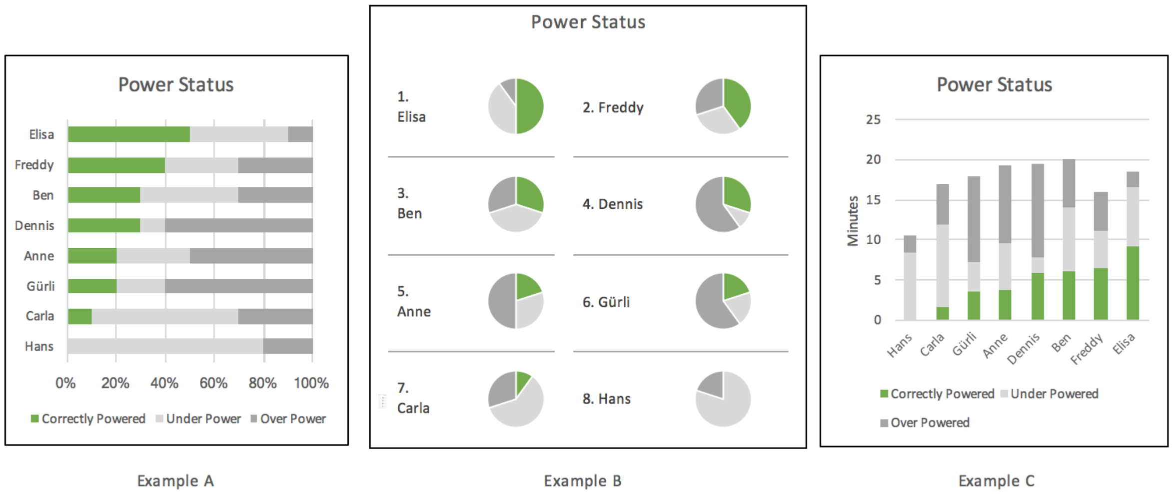


Figure 10: The three visualization variations for power status

The second set of visualizations, titled Median Time on Task, exemplifies three ways of communicating how much time students spend on a given task. The examples refer to data from the wind farm simulation were there are no predefined tasks as described in D2.2. Meaning the design of the labs is not designed towards tasks but rather a sandbox environment, making tracking more difficult. The task depicted refers to the time is takes for a student to stabilize the power after the power production is out of balance by either being over or under powered. Since this task is encountered repeatedly in the lab, the visualizations showed the median for the tasks. Example A in the set simply shows a list of the median time on task ordered from low to high. In example B the same is depicted in a horizontal bar chart. In example C the median time on task for the eight students is plotted on a chart with a dotted line showing the class median as additional information.

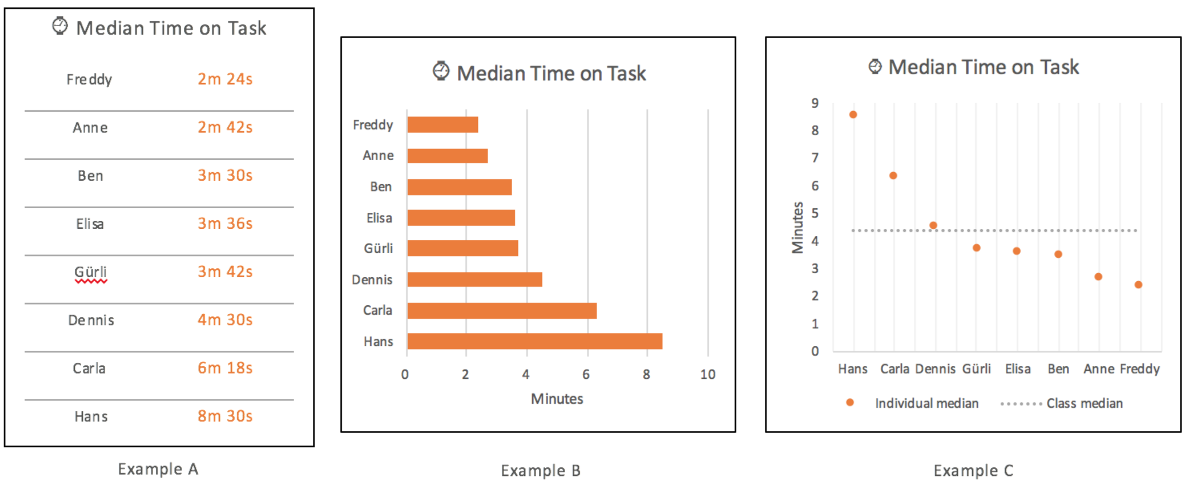
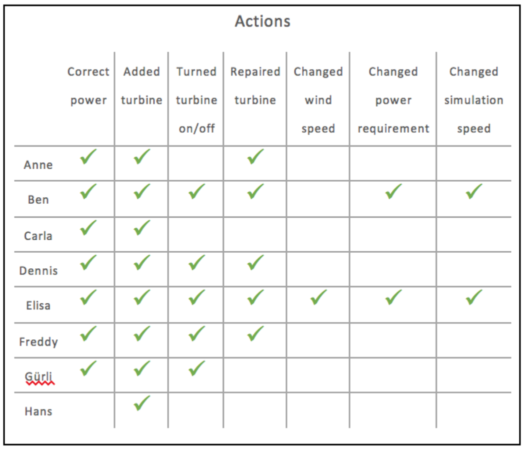


Figure 11: The three variations for time on task

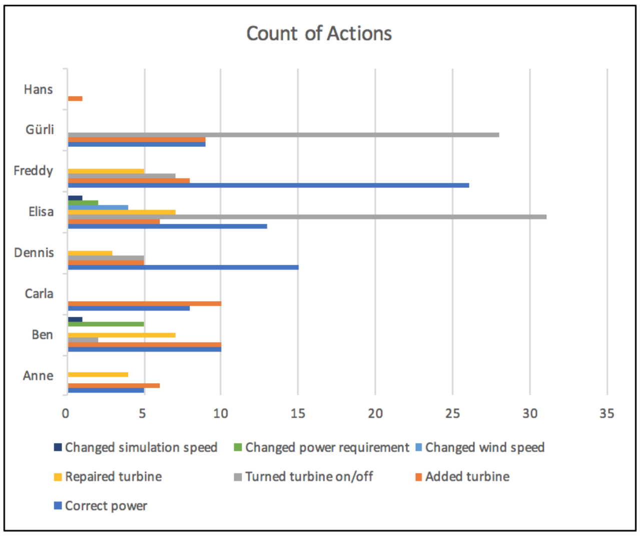
The last set of visualizations, titled Interactions, showed the actions the students took in the lab. The first simply shows, with check marks in table, which actions the students performed, in a different type of lab this type of visualization could be used to give an overview of the tasks students have completed. The second visualization is a horizontal chart. It depicts the same actions as the first but in this version the actions taking are counted. This last set of visualizations are quite different in their approach and potential use so in this case the testers were not asked to rank the two examples.



Actions

	Correct power	Added turbine	Turned turbine on/off	Repaired turbine	Changed wind speed	Changed power requirement	Changed simulation speed
Anne	✓	✓		✓			
Ben	✓	✓	✓	✓		✓	✓
Carla	✓	✓					
Dennis	✓	✓	✓	✓			
Elisa	✓	✓	✓	✓	✓	✓	✓
Freddy	✓	✓	✓	✓			
Gürli	✓	✓	✓				
Hans		✓					

Example A



Example B

Figure 12: The two interaction visualizations

3.4.2 Test participants

The test participants were the same as for the first test of the analytics tool visualizations, see Section 3.3.2 .

3.4.3 Results

Overall the test participants' responses to the visualizations were rather schematic with few elaborations. For the most part they simply linked the examples to the positive and negative terms mentioned without any further explanations or described the visualizations in laconic terms. This scarcity in feedback could be due to the phrasing of the text accompanying the visualizations and the fact that there was no facilitator present to ask in-depth follow-up questions. One could also speculate that, other teachers who teach different subjects might have given different and perhaps more verbose feedback.

In addition to the limited feedback, the test participants in few cases contradicted themselves. For the power status visualization one teacher listed example B with the pie charts as the least useful, least informative and as giving the worst overview but then ranked it as the best of the three. The contradictions make the part of the feedback difficult to use but there were comments that are valuable going forward.

Time on Task Visualizations

For the time on tasks example two of the tester commented positively on the line illustrating the class median on the plot chart, one of them even suggested adding it to the bar chart. If the teachers are interested in enriching the collected data this simple type of indication could also be used to illustrate for instance a performance measure for the task.

Also, as part of the comments for the time on task visualizations, one of the teachers notes that, depending on the number of students, it might differ which type of visualizations would give the best overview. With few students, one might be able to digest a more detailed table with the raw time data but for a larger number of students it would be beneficial with a more simplistic chart.

Comparing the bar chart and the plot chart for the time on task, it was noted that the two were very similar aside from the indication of the overall mean. One tester appreciated the finer granularity in the plot chart, a detail that in the final analytics front-end most likely would depend on the screen size. Another mentioned that the bars were better than the markers on the plot chart most likely because they are visually more striking.

Interaction Visualizations

For the interactions visualizations, the test participants agreed that, although the bar chart that depicted the count of interactions contained more information, the simply table gave a far better overview about which students did all the actions. One tester suggested that, since the bar chart contains more information, it could potentially be more valuable but he would need a further explanation about how to use the additional information. This point to the two use cases for the analytics visualizations in the ENVISAGE project one providing insights for the creators of the virtual labs and the other providing insight for the teachers using the virtual labs in the class room. The visualizations for the two use cases may at times overlap but might also require different types of visualizations for the same data.

Power Status Visualizations

For the stacked bar chart in the power status example, one tester noted that it was easy to compare the bottom columns between the students. This suggests that for this type of chart it might be harder to compare the data between students if there are many layers.

Based on the comments the pie chart in the power status, disregarding the ranking, all three testers thought it gave the worst overview of the three, was the least useful, and least informative. Perhaps this result was in part because the chart was missing the legend and had no numerical values connected visually to it.

3.4.4 Summary

One of the central takeaways from this evaluation of the different visualizations is that, deciding which visualization is the right one not only depends on the data being conveyed but also on the receiver. This is of course a basic tenet of communications but it is good to be reminded of this, since the purpose of analytics in ENVISAGE is twofold. While the metrics behind a visualization for the teacher who seeks to improve a virtual lab might be the same as the metrics behind a visualization for the teacher who seeks to gain insights into the students' performance in the class room, the two visualizations perhaps should not be identical. For instance, one might have the time to dwell on a more detailed visualization while the other might need a faster overview.

In the initial deliverable D1.1, the Greek teachers sought analytics that show the students categorized by performance. This is mirrored in the response from the teachers in this survey who liked the option of comparing the students to the class mean. Going forward it is worth to consider and test how similar tools for comparison can be incorporated into the analytical visualizations.

Another takeaway from this iteration of the analytical visualizations is the importance of having a facilitator present when the teachers evaluate the visualizations to ask follow-up questions and follow any additional avenues that may open and expand the feedback. For this extra iteration, this was unfortunately not possible, but for the next round of evaluations interviews should be conducted instead.

3.5 Virtual labs and learning content

3.5.1 Methodological approach

To evaluate the type of virtual labs and learning content that can be created using the authoring tool the teachers were asked to assess a demo version of such a lab, i.e., the Wind Farm Simulation [10]. Though this lab is not created using the authoring tool the Wind Farm Simulation exemplifies the type of learning content the teachers going forward will be able to make using the authoring tool.

The test participants were given a questionnaire and asked to consider, on a Likert-like scale, how strongly they agreed or disagreed with statements that cover subjects such as the students’ engagement with the lab, quality of educational contents, the fit in terms of the students’ abilities, and the teachers’ expectations.

Ideally the virtual labs are to be evaluated based on sessions with students and teachers using the virtual labs. However, due to time constraints for this first iteration of evaluations it was not possible to have the teachers test the virtual lab in a class room context. Instead the test participants evaluated the lab based on their experience as teachers and the original questionnaire presented in D5.1 was adapted to fit the test situation.

3.5.2 Test participants

The test participants evaluating the virtual lab were the same as for the user test for the authoring tool see section 3.2.2.

3.5.3 Analysis of Test Results

This section will summarize the teachers’ response. As mentioned before, 5 test participants is too small a sample to draw strong conclusions from but the feedback from the teacher was for the most part identical or very close.

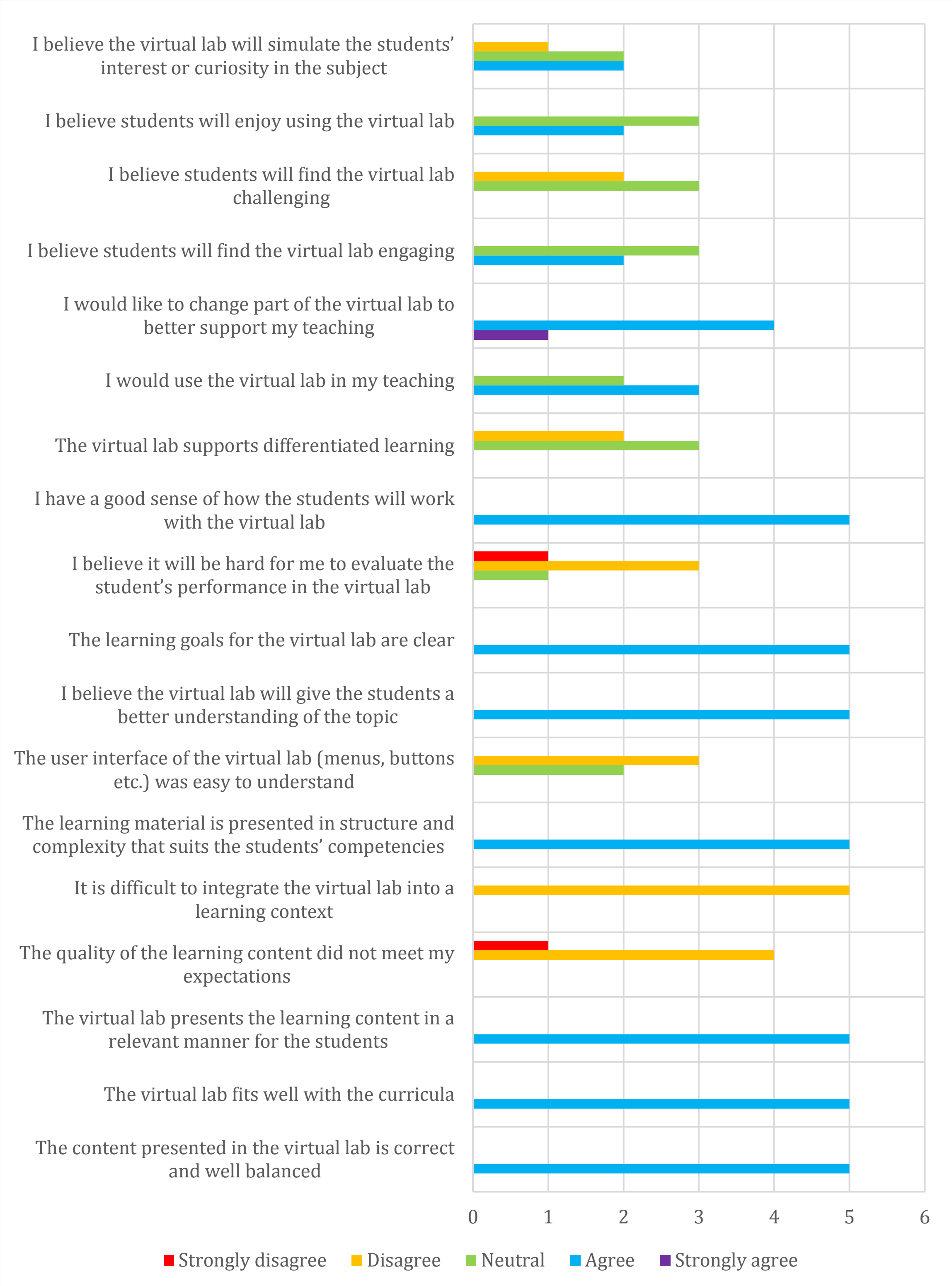


Figure 13 Summation of the teachers' response to the questionnaire on the virtual lab.

The chart above summarizes the teachers' responses to the statements in the questionnaire concerning the wind farm simulation used as demo lab for this part of the evaluation.

The following section will summarize the responses thematically.

Learning Content

With regards to the quality of the learning content and the fit with the curriculum, the teachers were unanimous in agreeing with the positive statements and disagreeing with the statements that reflected negative attitudes towards the virtual lab's ability to be integrated the into a learning context and their expectations of the learning material presented. As the tested lab for this first iteration of tests was only one of the ENVISAGE labs, there might very well be different opinions on the quality of the content of other labs but it is at least an indication that the wind farm simulation fulfils the teachers' expectations.

The Interface

The questionnaire had one statement related to the virtual lab's interface; "The user interface of the virtual lab (menus, buttons etc.) was easy to understand". Two testers were neutral while the remaining three disagreed with the statement thus stating that the interface is not easy to understand. The wind simulation is intended to be used by students in a class context so in addition to the teachers' feedback on this point it will be very valuable to get input from the students. Children and teenagers may have a very different experiences and expectations of digital interfaces compared to adults. Going forward this should also be assessed.

Perception of Students' Attitude

The statements relating to the students' engagement and enjoyment of the virtual lab, as well as the ones about how challenging and simulating the virtual lab was, were originally intended to reflect the teachers' observation of the students interacting with the lab. Without the students as a part of the initial evaluation the feedback on students' attitudes is naturally hard to give, something that seems to be reflected in the answers given by the teachers. The answers in this category are concentrated on the neutral midpoint.

Authoring and Differential Learning

Though the teachers were satisfied with the content in the lab they all agree they would want to change part of the lab to better support their teaching. This fits well with the ENVISAGE goal for the authoring tool, a goal the teachers were aware of, since they also tested the authoring tool.

One of the questionnaire statements referred to whether the virtual lab supported differential learning, a concept that the project going forwarded aspires to incorporate. The teachers either disagree or were neutral in their response which also matches with the demo version they were asked to evaluate, as this does not yet have adaptive features imbedded.

3.5.4 Summary

Overall, teachers evaluated the virtual lab in a positive manner. They were happy with the content presented in the Wind Farm Simulator, they believed the lab would give the students a better understanding of the subject and they saw the learning goal for the lab as clearly defined.

It is naturally important to note that the evaluated lab only represents one example of a lab that could be created through the authoring tool. For different virtual labs, the educational content would of course need to be evaluated on its own.

The user-friendliness of the lab was a part of the evaluation by which the testers appeared to be less impressed. As mentioned above, this is where the feedback from students is relevant to take into consideration for future tests as the teachers, as well as the students, are the intended end-users and a younger audience might have different expectations and requirements when it comes to, for instance, user-interfaces.

Fitting with the goals of ENVISAGE the teacher indicated an interest in being able to change parts of the lab.

Lastly, this test was conducted with a demo lab that was not created using the authoring tool but rather a lab exemplifying a lab that could be made with the authoring tool. For future tests it will be necessary to evaluate labs that follow the ENVISAGE ecosystem and are created using the authoring tool.

4 Summary and conclusion

In total five pilot were conducted for the first iteration of the ENVISAGE services. Two for the authoring and visualizations tools and one for the virtual labs. For the first pilot of the authoring tool (heuristic evaluation), a general lack of helping functions and tooltips was found. Furthermore, the authoring tool was also using ambiguous term, which could easily mislead the user. During the second pilot (user testing), the teachers struggled when using the authoring tool and reported that they would probably not use it before it became easier for them to use. However, they saw a great potential in using the system and they are hence potential users of the finished product. Improving the UI further would therefore help accommodating the user's requests and hopefully, attract them as end-users. Encountering difficulties in the early version of an IT-system is to be expected, however the participants still expressed enjoyment about the possibility of being able to create a 3D experience for their classroom. The overall idea of the tool is therefore not unfamiliar to the possible end-users and the results support the concept as a whole. To obtain a higher degree of usability for the authoring tool, the UI would have to foster more user-friendly control of the 3D environment, as this was the participant's biggest concern during the second iteration, a valuable finding going forward.

For the first pilot of the analytics and visualizations tool, the participants found a substantial amount of the visualizations hard to read and apply. The other half of the visualizations had a more basic nature and the participants had encountered them before in different contexts. Because of this, the participants intuitively knew how to read them based on previous experience. The conclusion from this initial pilot was that the visualization had to be more simple and basic then first anticipated by the researchers. A second pilot using a questionnaire as method enabled a focus on more simple visualizations, which were iterated in response to findings from the first pilot, and pinpointed which could most efficiently communicate a specific metric. A central takeaway from the second pilot is that the appropriateness of a visualization depends on not only the data being visualized but also the receiver hereof. Overall, the visualization from the second iteration appears to be easier to read for participants, as none of them reported not being able to read the visualization presented for the second pilot.

Overall, the test participants evaluated the labs with positive feedback. They liked the content and believed the wind energy lab would give the students a deeper understanding of the subject. The also clearly saw the learning goal of the lab. The more negative comments were connected to user-friendliness of the lab, however the teachers are not the direct end-users of the lab and students opinion on the labs would thus here be more valid. It is also crucial to note that this was merely the most negative element evaluated and the degree of negativity did not appear threatening to the overall experience of the lab. Also for the virtual labs, the teachers showed a great interest in being able to change part of the lab.

Generally, the users saw great potential for both the analytics and authoring tools but also for the quality of the labs that can be produced. However, they also noted issues and errors which need to be corrected before the services within ENVISAGE is ready to be used in a real-life context. This is not unusual feedback when developing new and innovative IT-solutions and strategies for integrating and responding to the feedback have already been planned for in the future WP's of the project. The next six months of the project is therefore used for redesigning the solutions based on the feedback we gained through this iteration and subsequently conduct more pilot studies to evaluate if the quality again during the second iteration. The quality of the ENVISAGE tool has been seen to increase with every cycle of pilot, insights collection and changes.

In future evaluations, additional end-user could also be the focus of the tests. For this iteration, only teachers from the same school Ellinogermaniki Agogi have been used as participants. As Ellinogermaniki Agogi is a partner of the consortium, using the Greek school as the primary source and context for pilots is only natural but cultural differences might be visible if the solutions are tested on participants from another country. Another direction that the future evaluations could take is expanding the test for the additional end-user previously identified. Developers of virtual labs, games or similar applications for education could thus also generate more insights for the development and enable the ENVISAGE services to be used more broadly than just teachers.

5 References

[1] E. Consortium, Grant Areement, European Commission, 2016.

[2] J. Nielsen, Usability engineering, Philadelphia, USA: Elsevier, 1994.

[3] “Virtual labs authoring tool, Help,” [Online]. Available: http://augreal.mklab.iti.gr/envisage/authortool/help/.

[4] J. M. Billson, Focus Groups: A Practical Guide for Applied Research,, Clinical Sociology Review: Vol. 7: Iss. 1, Article 24. , 1989.

[5] C. Holmgård and F. Hadiji, D2.3 Visualization Strategies for Course Progress Reports, In ENVISAGE: , 2017.

[6] A. J. Pickard, Research Methods in Information, vol. II, London: Facet Publishing, 2013, pp. 243-249.

[7] K. S. Bordens and B. B. Abbott, Research Design and Methods A Process Approach, New York, NY 10020: McGraw-Hill, 2011.

[8] R. Likert, A technique for the measurement of attitudes, New York: Archives of psychology, 1932.

[9] G. Mavromanolakis, D5.1 Pilot execution plan and evaluation protocol, In ENVISAGE , 2017.

[10] ENVISAGE, “Wind Farm Simulation,” 2017. [Online]. Available: http://www.envisage-h2020.eu/games/energy/v1_1_3/.

[11] F. D. Davis, “Perceived Ease of Use, and User Acceptance of Information Technology,” *Management Information Systems Quarterly,* pp. 319-340, 1989.

[12] J. R. Lewis, “Psychometric Evaluation and Instructions for Use,” *International Journal of Human-Computer Interaction,* pp. 57-78, 1995.

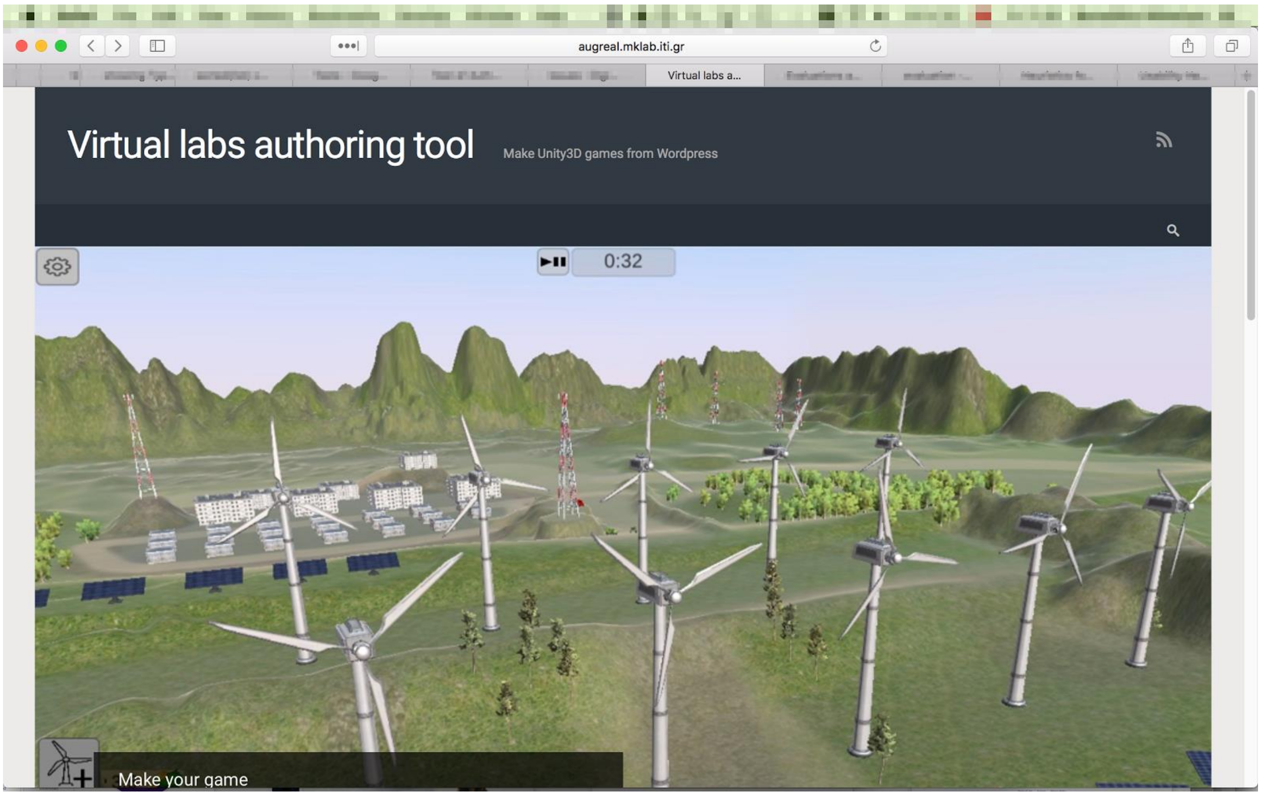
[13] J. Nielsen, "nngroup.com," nngroup, Oct 2010. [Online]. Available: https://www.nngroup.com/articles/mental-models/. [Accessed 22 Oct 2017].

6 Appendix

6.1 Usability Report: A heuristic evaluation of the authoring tool

6.1.1 Login page

The login page is where the user login to the authoring tool. Here you can also find information more information about the project and the functionalities of the authoring tool.



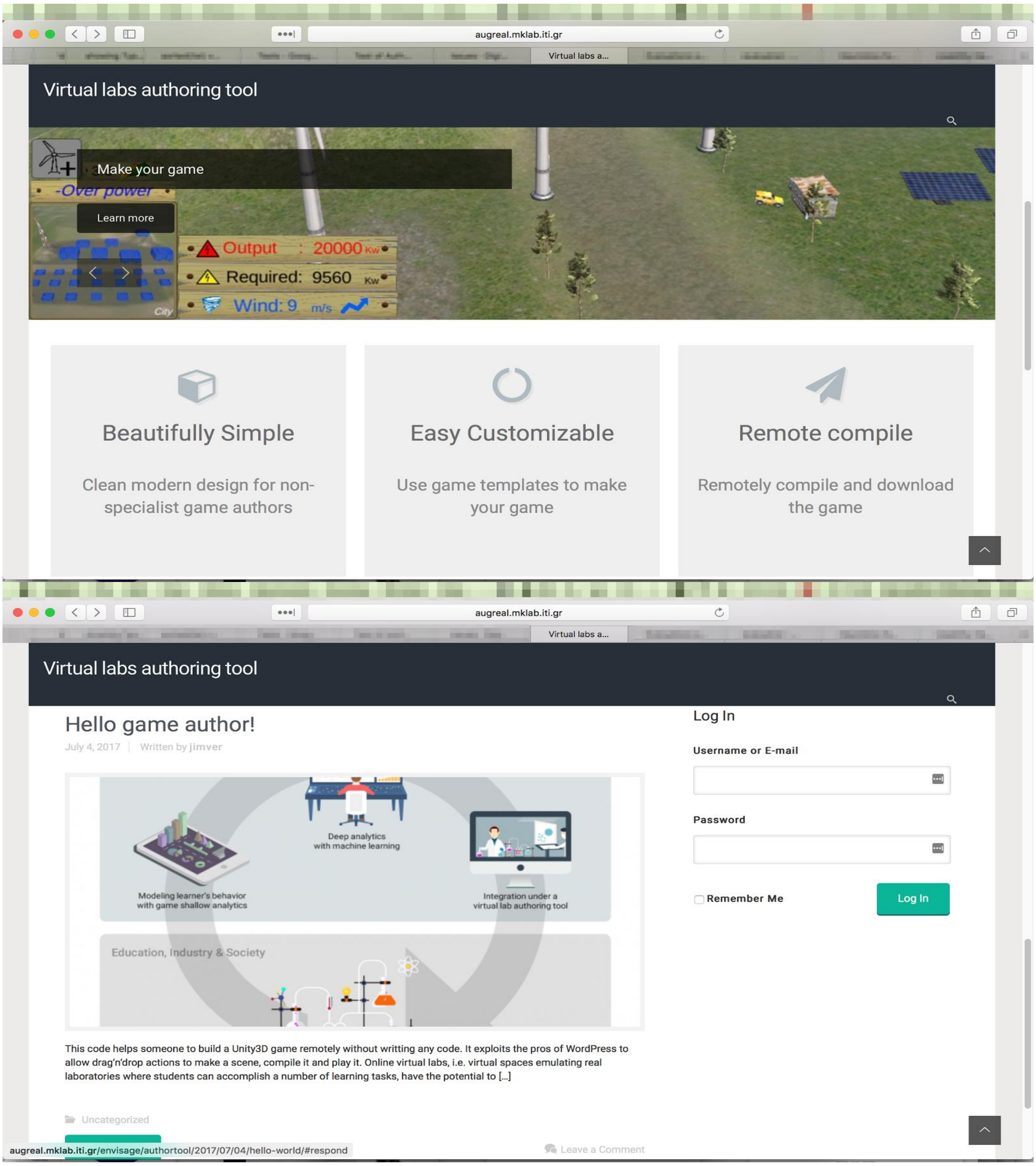


Image 1: Screenshot of the login page.

Consistency and standards

- Consistency in terminology: The authoring tool switches between using the term “Game” and “Virtual Lab”(see image 2 above).
  Recommendation: For consistency only use one term e.g. Virtual Lab

Flexibility and efficiency of use

- The login field is placed at the end of the page below the fold. This makes it harder to find for a first time user and is inefficient for the returning user.

Recommendation: Move the login field to the top part of the page e.g. on the banner image or in the top bar.

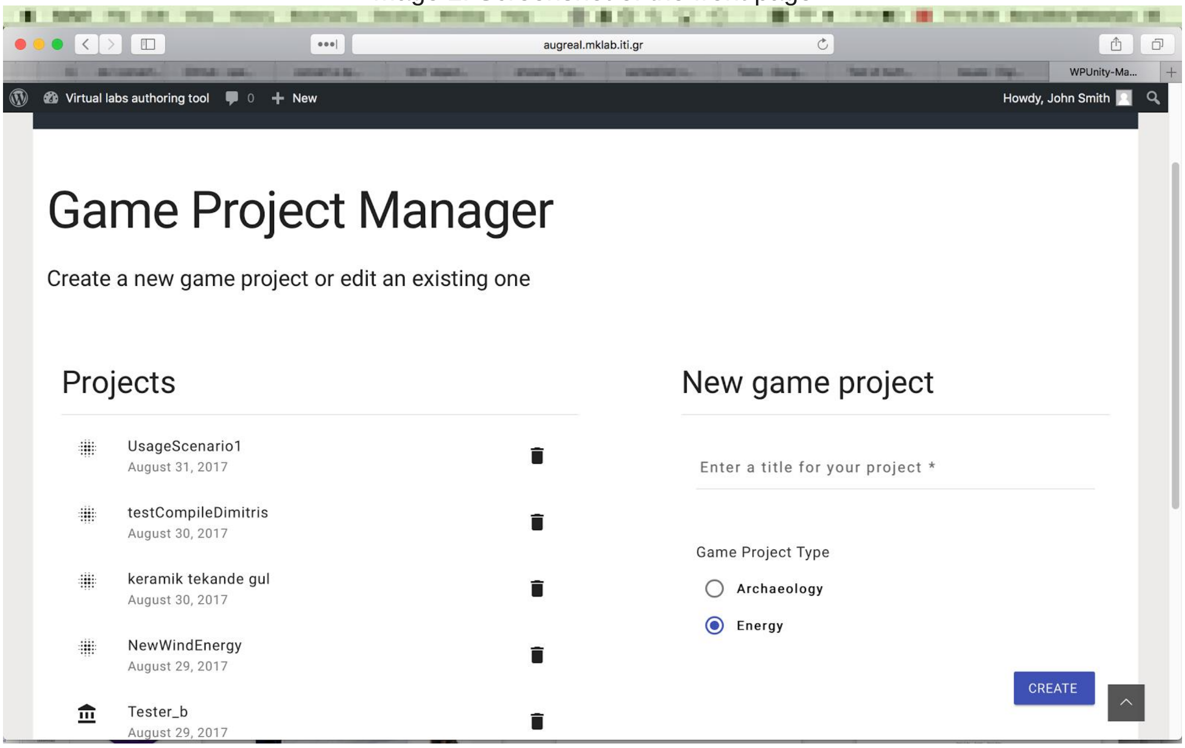
Aesthetic and minimalist design

- The buttons (“Make your game”, forward and back) on top of banner image do not link to anything.
  Recommendation: Remove these or add the relevant links so they match a user’s expectations.
- The RSS feed button is irrelevant for the authoring tool use case.
  Recommendation: Remove it.
- The comment functionality and the search field seems irrelevant for immediate use case for the authoring tool.
  Recommendation: Consider the use case for these functionalities.
- The banner image depicts the interface of the virtual lab including the buttons. This is confusing to a first time user.
  Recommendation: consider cropping the image or use an image that places the virtual lab in a context for instance on a screen with a student in front.

6.1.2 Front page (Home)

The front page is the landing page the users are sent to after logging in to the authoring tool. Here the users can create new projects or continue working on existing.

Image 2: Screenshot of the front page.



Consistency and standards

- Consistency in terminology: The functionality of “Projects” could be unclear to new users as the list will most likely be empty.
  Recommendation: Making the title more active and changing “Projects” into “Existing projects” or similar could help explain the functionality of the list better.

- Consistency in terminology: The functionality of “New game project” could be unclear to new users.
  Recommendation: Making the title more active and changing “‘New game project” into “Create new project” or similar could help explain the functionality of the list better. This would also provide more consistency with the “Create” button for the function.

Match between system and the real world

- Active word use: “Game project type” could be confusing for users, as the authoring tool is for creating virtual labs and subsequently because it does not inform the user of what action is required to continue.
  Recommendation: Making the title more active by changing “Game project type” into “Choose project type” or similar could help explain the action required by the user more.

Help and documentation

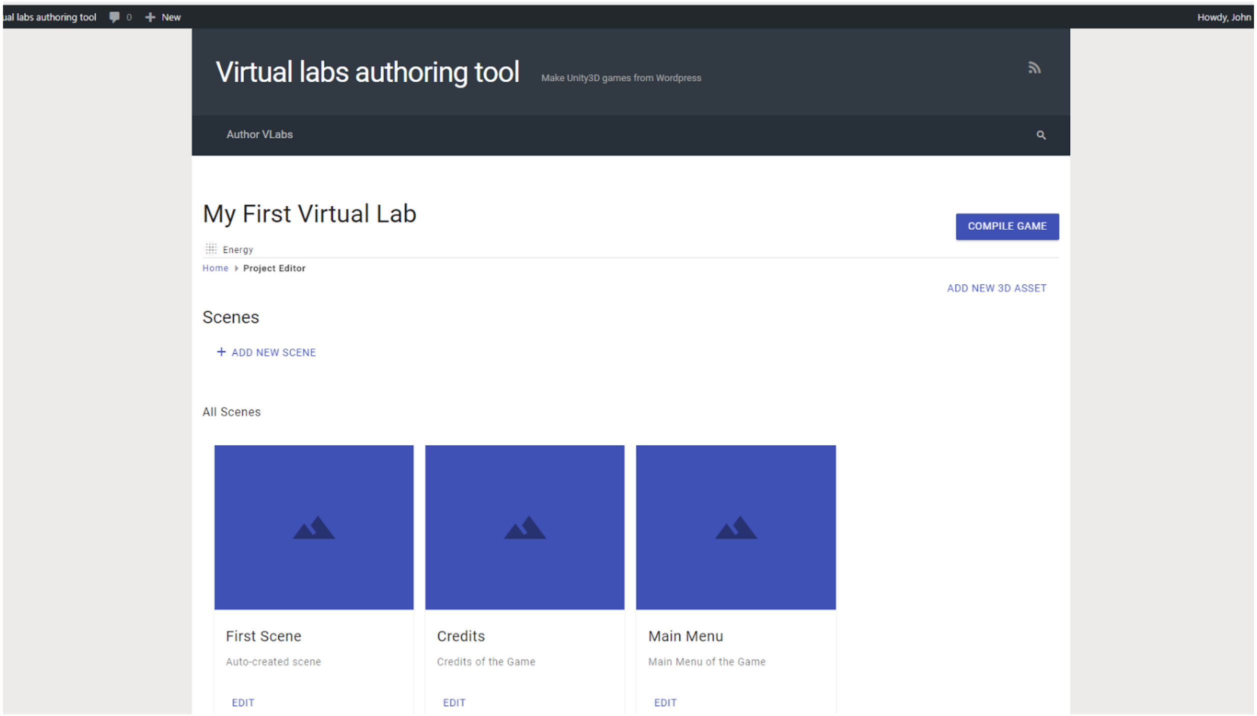
- Documentation needed: The two project type are not explained further and no additional information about them is offered to the user.
  Recommendation: Explain the project types to the user by either having a “more details button”, link to an additional page explaining the types or similar.

Aesthetic and minimalist design

- The dark bar below the wordpress topbar consists of a header that links to the login page and an unused menu bar with a search function. These elements are in their current state of little use to a user and take of space moving the actual content further down the page.
  Recommendation: Remove the dark bar or minimize it and incorporate navigation that embraces the relevant pages in the authoring tool.
- Unnecessary limitations: The title of the project can has a minimum of five characters for naming the project. While this might have an underlying reasoning behind it, it does not make sense from a user experience point of view, as it limits the user.
  Recommendation: Remove the limitations on characters to allow the user more freedom when choosing the name of a project or scene (This goes for the whole authoring tool)

6.1.3 Project editor

After creating a new or entering an existing project, the user is sent to the project editor. On this page the user can create new 3D scenes or asset, compile or edit an existing 2D or 3D scene.



Help and documentation

- The project editor presents an overview of the scenes that in the end will make the finished virtual lab but there is no explanation neither visual or textwise of how the different elements are connected in the compiled virtual lab. Leaving the user with questions like: When are the credits shown? In the case where the user creates multiple scene in which order will these appear in?
  Recommendation: Adding explanatory texts and previews even if they are only simplistic would greatly help guide the user.
- System terminology: Unity terms like “Asset”, “Scenes” and “Compile” may well be unfamiliar to the user.
  Recommendation: Add explanatory text to or consider using more self explanatory terms.

Match between system and the real world

- The structure of the workflow when creating a lab: The first elements the user can interact with on the page are the “Compile Game” button, the “Add new 3D asset” and the “Add new scene” links. However when looking at the ideal workflow for the users, this might not be the ideal first interactions to start with e.g. compiling is most like the last. When the user first enters the project editor there are three predefined scene already set up. On a smaller screen these are displayed below the fold as these represent the minimal lab project a user can create. Maybe the default scenes should be moved further up on the page to match the minimale workflow.
- Recommendation: Think about the ideal workflow for the user and let the design facilitate this. Structuring the elements in an order that match the ideal workflow for using the authoring tool would also improve the usability.

Consistency and standards

- Navigation: With the dark header bar at top of the page pushing the page content down the breadcrumb trail is placed further down than conventions dictate making it

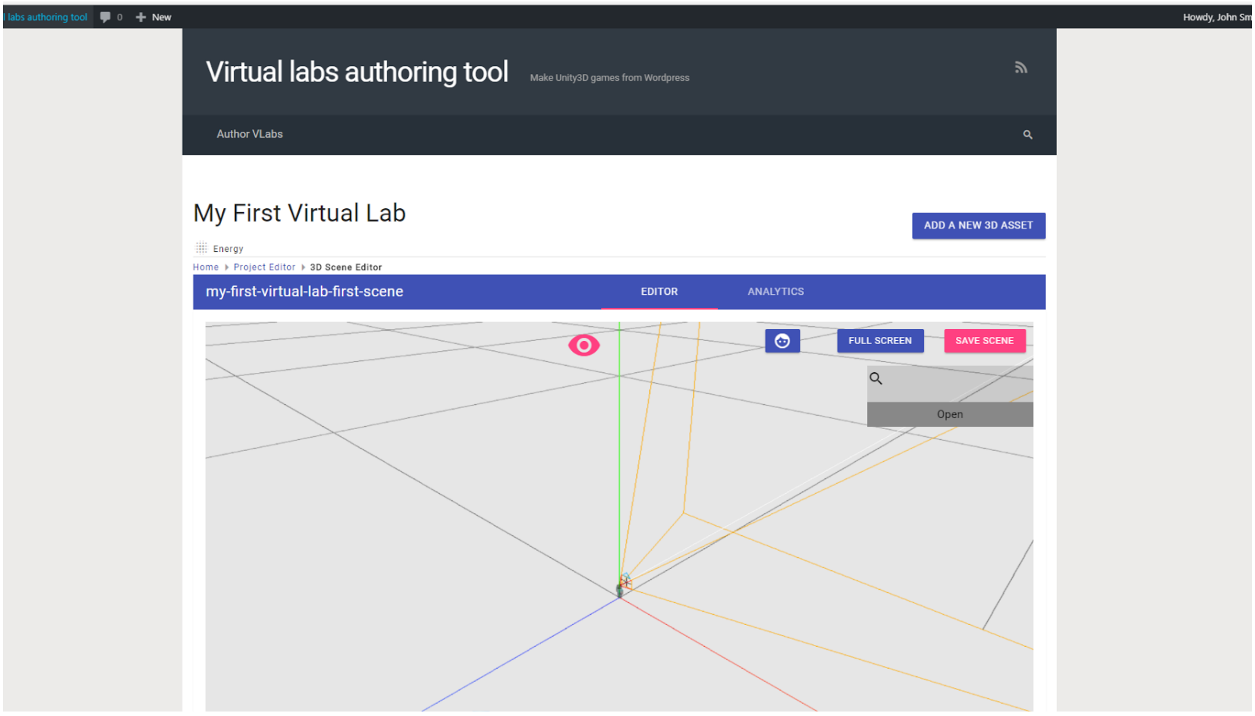
harder to find.
Recommendation: If the dark header bar is kept in place then move the breadcrumb further up.

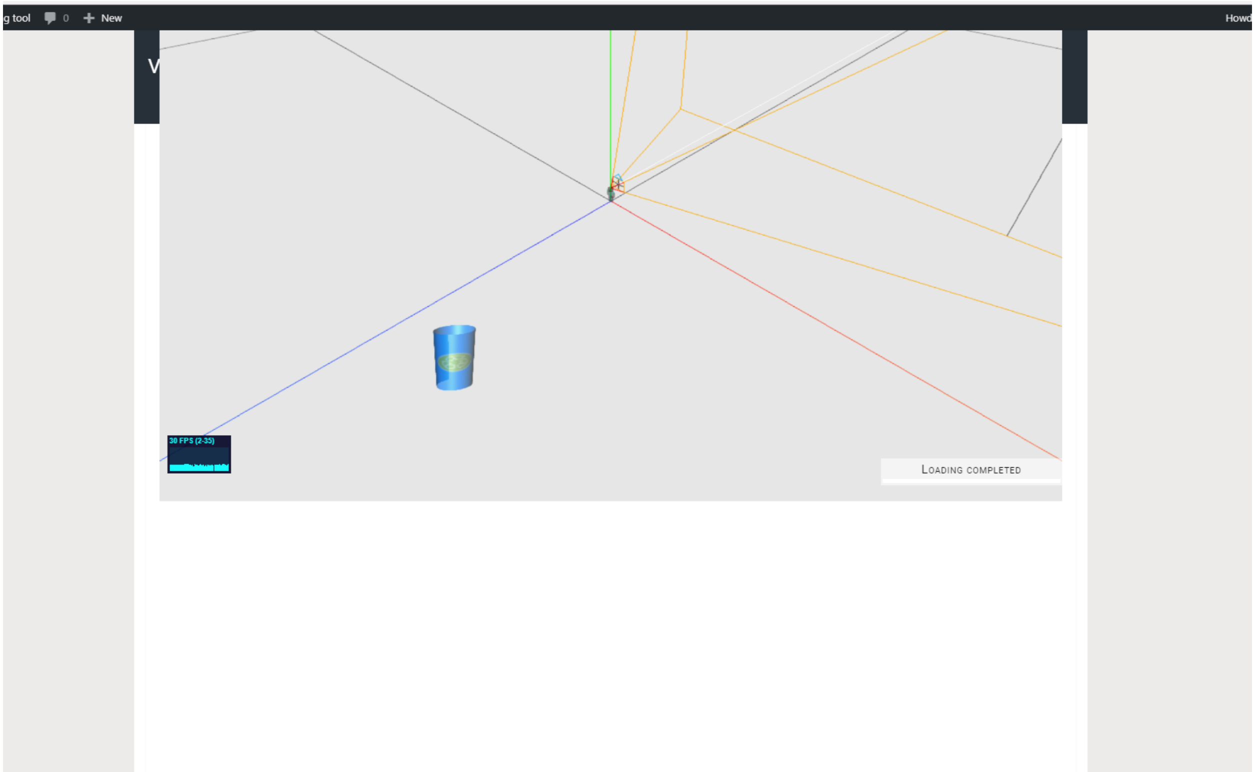
Visibility of system status

- It is possible to access the Add New 3D Asset functionality from the Project Editor via "Add new 3D asset" however once a user has gone through all the steps required and created an asset the user is returned to the Project Editor page without any feedback on where the asset is placed.
  Recommendation: Either remove the feature for adding assets on the page (it is already possible to add asset from the 3D scenes) or show a library of the created assets on the Project Editor page alike to the All Scenes functionality.

6.1.4 3D scene page

When creating a new or entering an existing 3D scene the user is sent to the 3D scene page. On this page, the user can add and move 3D assets around and make simple manipulations to the assets, essentially, setting the scene of the game.





Help and documentation

- Inadequate help: Though there are some tooltips on the interface in the 3D scene, they do not adequately explain the functionality to the user e.g. the toggle button with an image of 2 arrows has the tooltip “Toggle double sided object”.
  There are also elements with no explanations such as the frames per seconds counter in the bottom left of the 3D view and the blue cylinder object that is fixed to the bottom of the view.
  Recommendation: Provide help and documentation that explains the features while avoiding technical language and in line with Nielsen’s heuristics uses phrases and concepts familiar to the user.

User control and freedom

- The user cannot undo and redo actions made in the 3D scene.
  Recommendation: Add redo and undo functionality
- On a smaller screen it is not possible to see the entire 3D scene at once and since scroll is used to navigate the 3D space it is extra difficult for users to grasp the scene in its entirety.
  Recommendation: Implement the 3D view as a responsive element that adjust to different screen sizes.
- A user can add a new asset using the “Add new 3D asset” button on the 3D scene page however once a user has gone through all the steps required and created an asset the user is sent to the Project Editor page and not returned to the scene where the user originally started this may confuse the user.
  Recommendation: Return the user to the 3D scene where they started or only keep the add asset button on the project editor page.

Error prevention

- There is a save feature in the 3D scene however there is nothing reminding the user to use it before navigating to other parts of the website. So a user may easily lose their work if they navigate away without saving.
  Recommendation: Add a warning popup or similar reminding the user to save before navigating elsewhere.

6.1.5 Credits

When entering Credits the user is sent to the 2D scene Credits. Here, the user can write a text that will be shown as the lab credits.

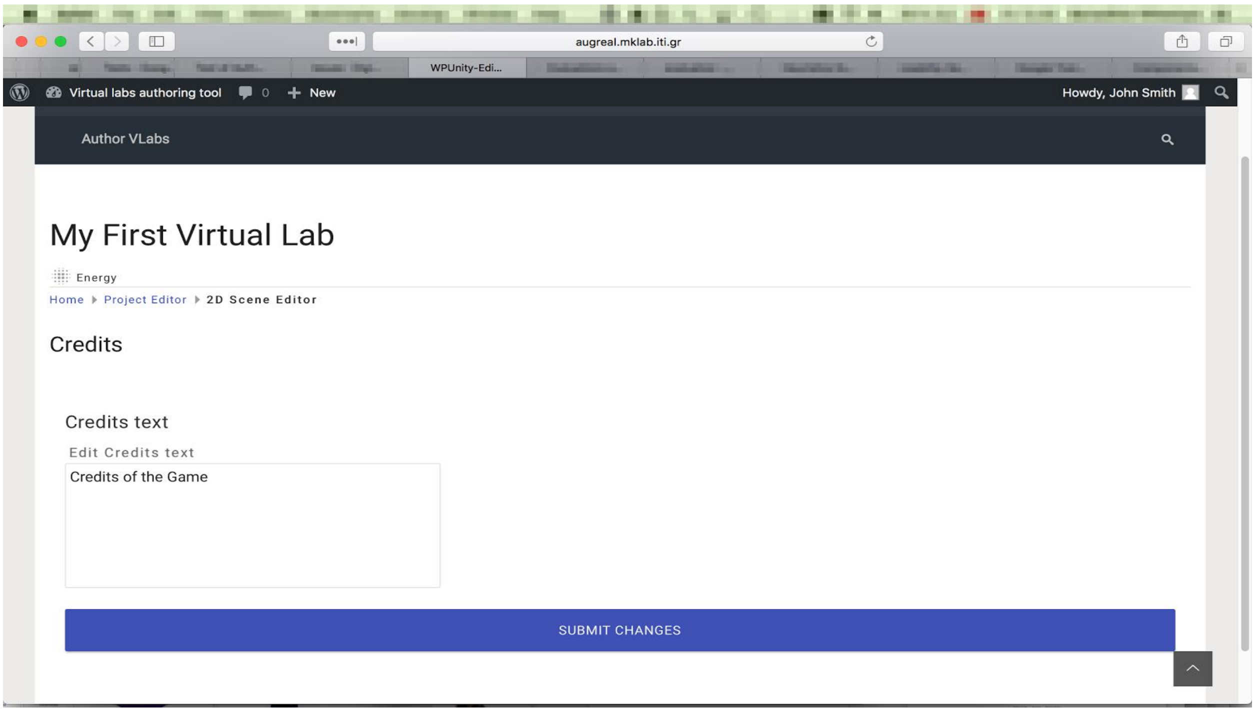


Image 5: Screenshot of the credits page.

Help and documentation

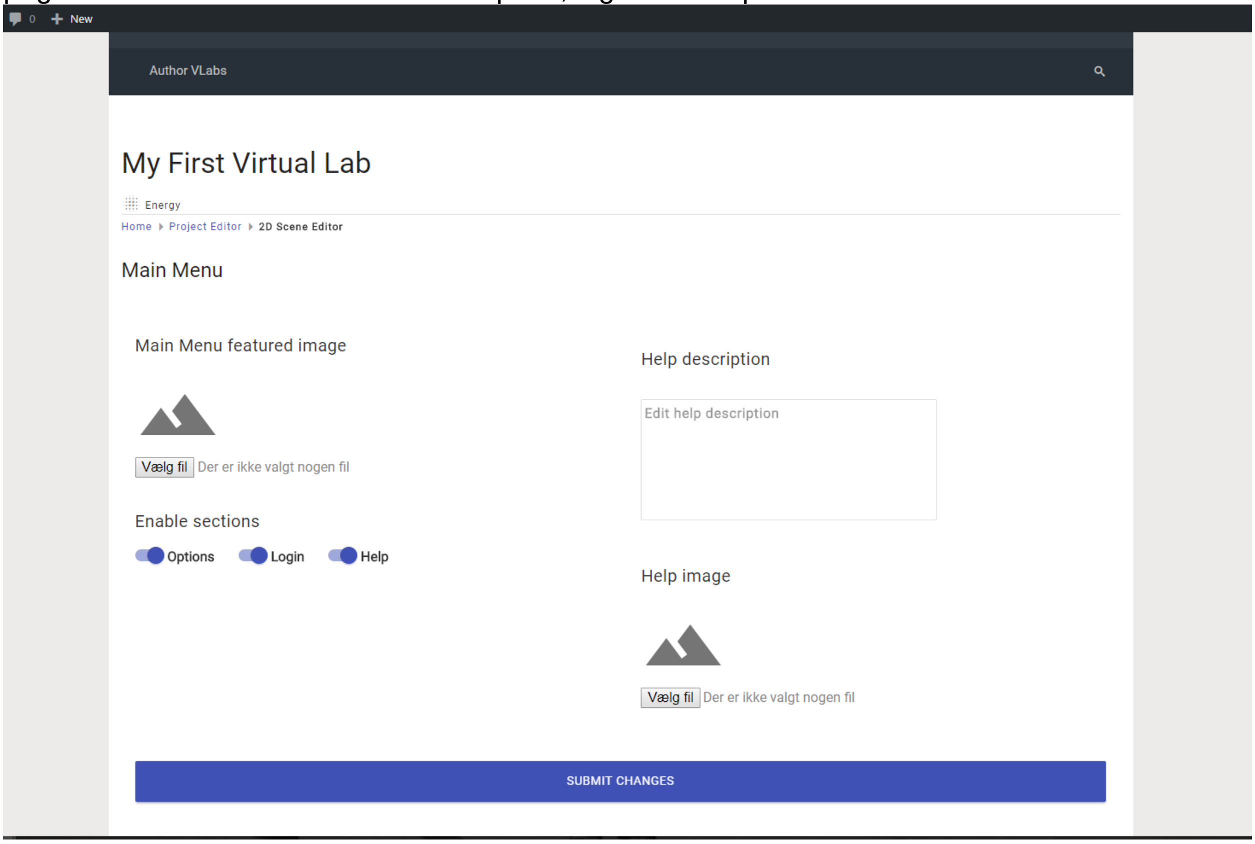
- Missing preview: The credits page only consists of a textbox and a submit button and the user is therefore not aware of where, how, when and it what format the credits will be shown.
  Recommendation: A preview of the credits would help document for the user how the credits will look.
- Description needed: The text describing the textbox says “Edit Credits text” and is therefore not offering the user any suggestions for what type of information the labs credits could contain.
  Recommendation: “Edit Credits text” could e.g. be removed and a text containing a more thorough description could be added below the header “Credits”.

Match between system and the real world

- Inconsistency in terminology: The breadcrumb trail for credits references the credit pages as a “2D editor” and while this might be a 2D scene editor, the terminology is more focused on a system thinking than the user-friendly terminology.
  Recommendation: Changes the name to Credits

6.1.6 Main Menu

The Main Menu page is entered when clicking the pre-existing scene in the Project Editor. On this page, the user can upload files as background pictures for the help and main menu page or en- or disable the functions: option, login and help within the lab.



Match between system and the real world

- Inconsistency in terminology: Main menu featured image can be an ambiguous term Recommendation: Changes the name to e.g. “Choose main menu background” to emphasis what the image will be used for.
- Inconsistency in terminology: The breadcrumb trail for credits references the credit pages as a “2D editor” and while this might be a 2D scene editor, the terminology is more focused on a system thinking than the user-friendly terminology. Recommendation: Changes the name to Main menu.

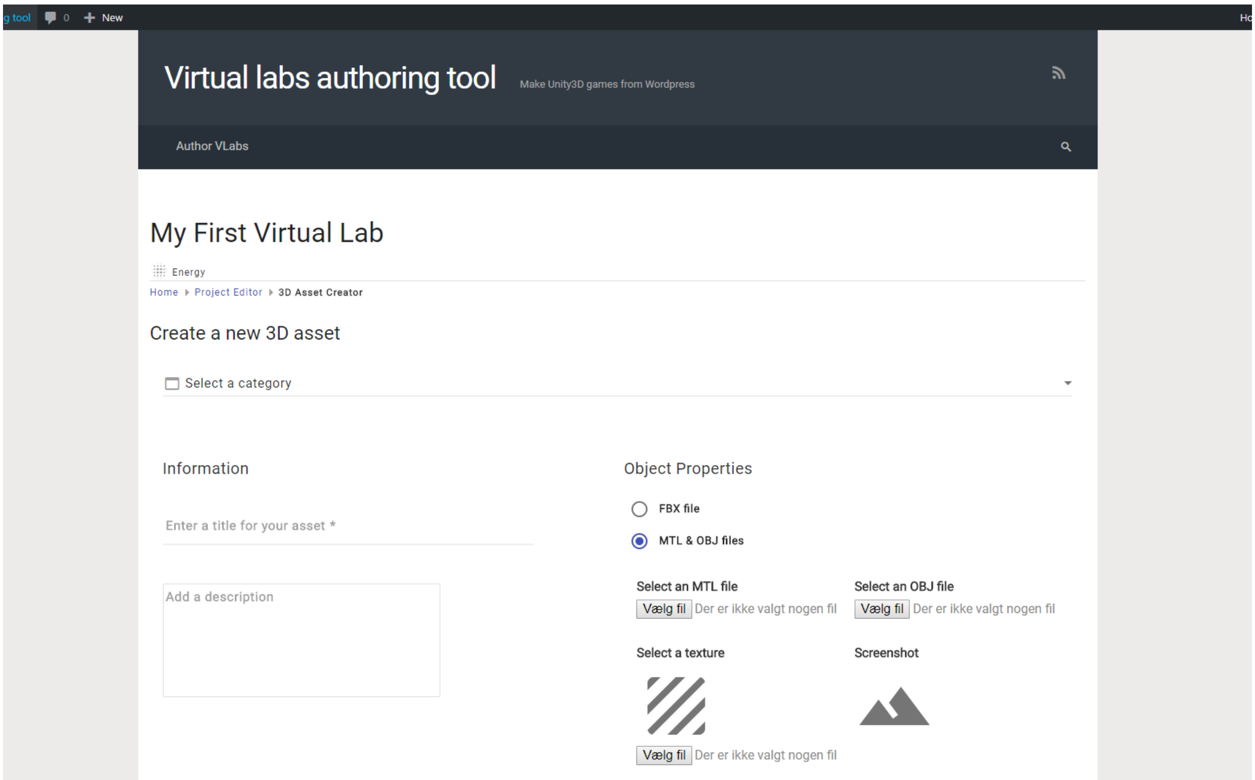
Help and documentation

- Missing preview: Although there is a preview of the picture (if you upload one), it is unsure what enabling the sections: “options”, Login and “help” will look like. (The Help pages suffers the same faith)
  Recommendation: have a preview where the user in real time can see the changes or a sketch explaining (not real time) where the different features and images will be placed.
- Missing preview: The Help pages suffers the same faith as above and it is also not clear if the help page is a page on its own or a part of the main menu
  Recommendation: have a preview where the user in real time can see the changes or a sketch explaining (not real time) where the different features and images will be placed.

- Missing description: “Enable Sections” are not self-explanatory, especially not Options, which could be a variety of things.
  Recommendation: Maybe a preview of the main menu pages (as explained above) would solve this issues and otherwise a tooltip, more information link/fold out or explanatory text could also help to improve this.

6.1.7 3D Asset Creator

The 3D assets creator is entered either through the Project Editor or the 3D scene page. Here, the user can add consumers, decorations, producers and terrains that subsequently can be used in the 3D scenes.



Error prevention

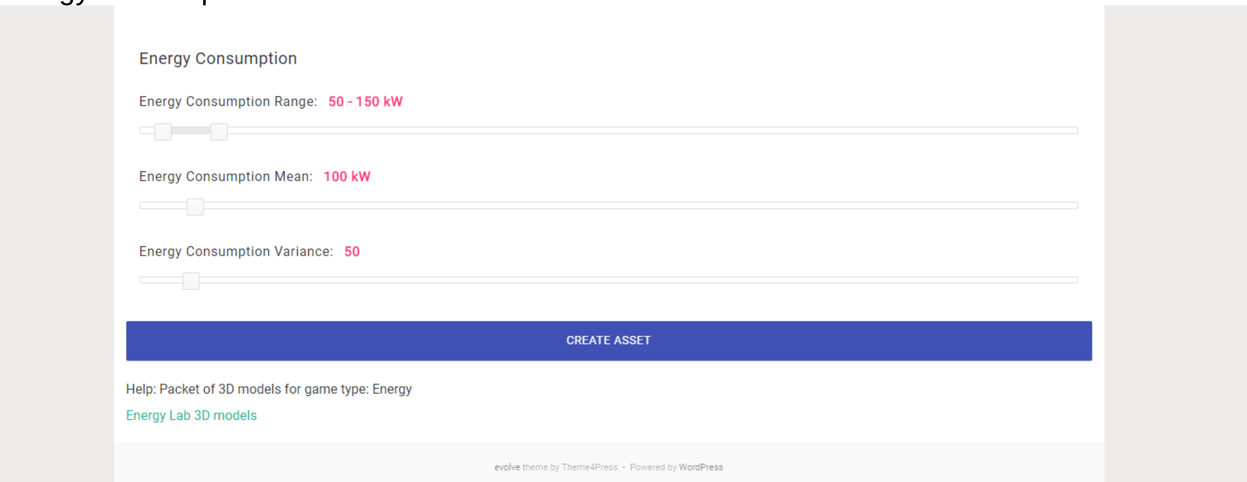
- It is easy to overlook the “Select a category” drop down and if a user first uploads an asset and then later chooses a category then the uploaded disappears.
  Recommendation: The Information and Object properties element should first appear once the user has chosen the relevant asset category.

Help and documentation

- Unclear description: It is unclear as a user where and by whom the text in the description box will be visible. In the scene for the users or in the 3D scenes? And why is it necessary?
  Recommendation: Maybe prolong the default text in the box explaining to the users where the text will be visible.
- Developers language: Maybe all users will not be aware of what FBX, MTL and OBJ files are and when you use one or the other. The same goes for textures.
  Recommendations: Offer the user explanations for the different types of files and when they are often used.

6.1.8 3D Asset Creator - Consumer

When choosing the category Consumer for a new 3D asset, the user will have to pick the energy consumption for the asset.

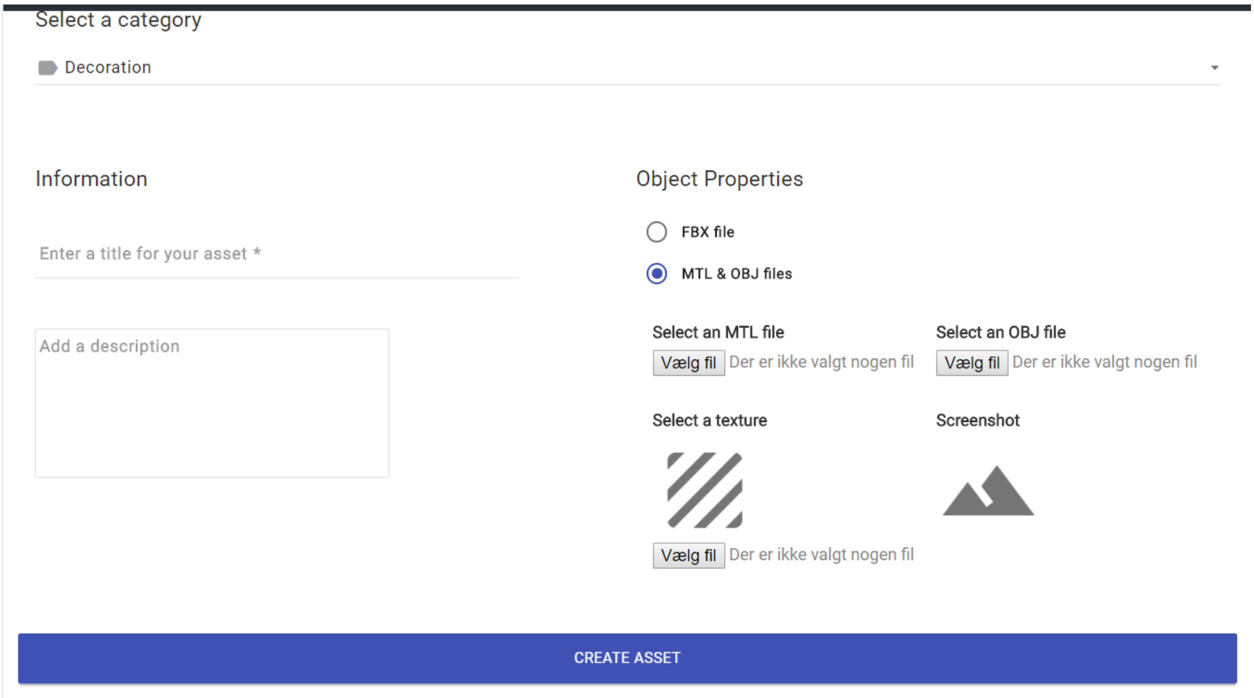


Match between system and the real world

- Double meaning: The word consumer can have a different meaning depending on the context it is used in and the user might be confused or mislead by the term. Recommendation: Supply the user with a bit of explanatory text when choosing the category. Some tooltips directly on the different parameter of energy could also guide the user's choice when setting them (e.g. mark by using a question mark)

6.1.9 3D Asset Creator - Decoration

When choosing the category decoration for a new 3D asset, the user can add asset that only have a decorative purpose and do that not directly affect the game.

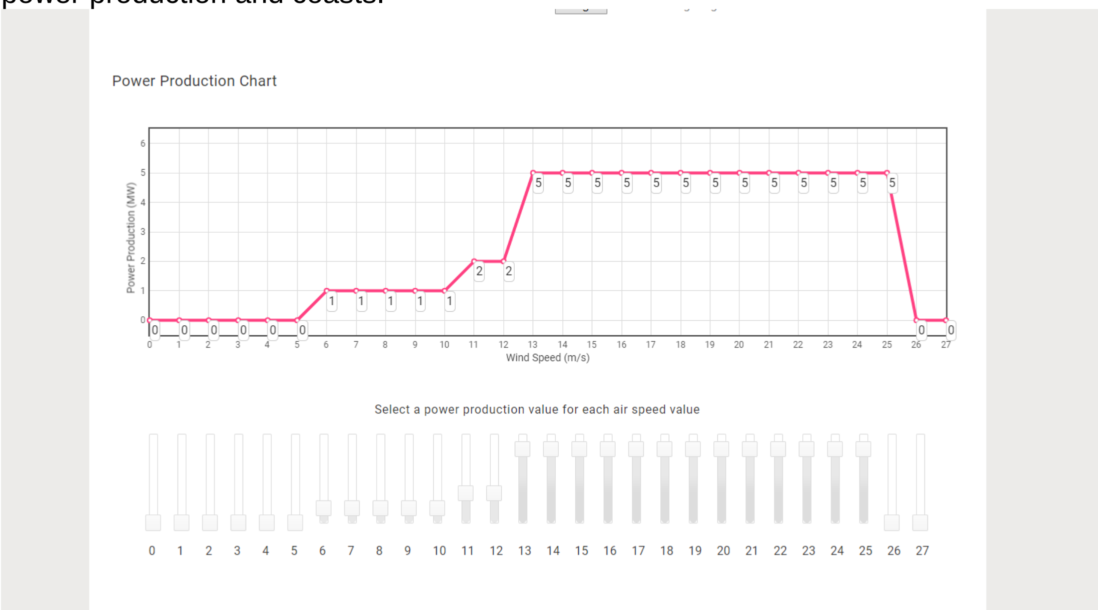


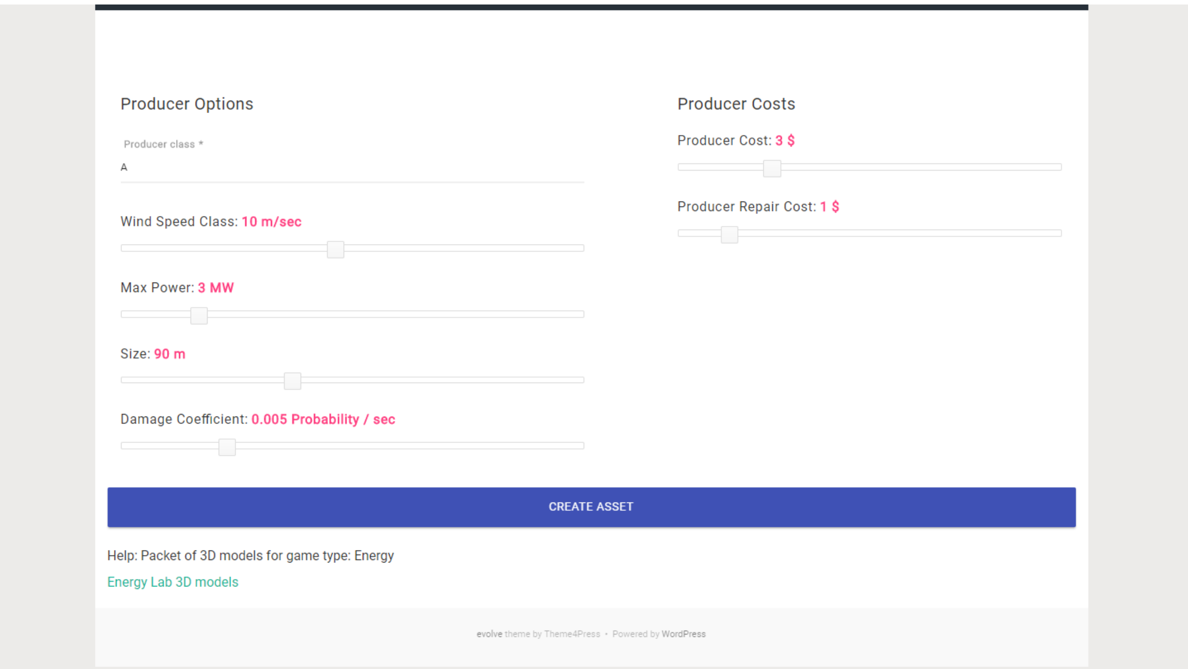
Help and documentation

- Missing description: the assets has no description.
  Recommendation: while the functionalities of a decoration asset could be more obvious to the users than e.g. consumer and producer, a text explaining this category’s purpose would reinsure or inform users further.

6.1.10 3D Asset Creator - Producer

When choosing the category producer for a new 3D asset, the user will have to select the power production and coasts.



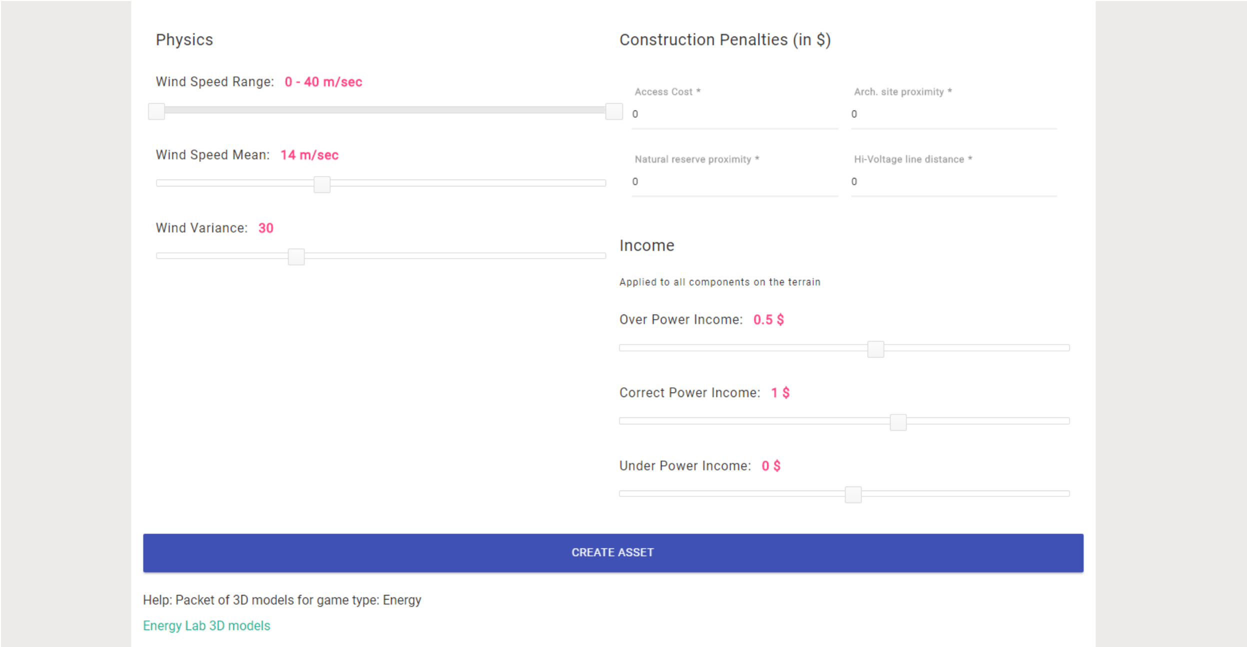


Help and documentation

- Missing documentation: When testing this asset, we became aware that this asset could be conceived as not being a 3D asset but settings for the wind range during the game,
  Recommendation: Explain to the user what the x and Y axis means, to ensure they make informed decisions when setting them.
- Missing documentation: The same issue as above is applicable for the producer options.
  Recommendation: Maybe tooltips or other explanations of the different settings to ensure they make informed decisions when setting them. Some decision made for one parameter might also affect the power productions charges or other producer options and the user needs to be made aware of this.

6.1.11 3D Asset Creator - Terrain

When choosing the category Terrain for a new 3D asset, the user will have to pick the physics and constructions penalties.



Help and documentation

- Missing description: there are settings Physics, Construction Penalties and Income. Recommendation: a text explaining this category's purpose would reinsure or inform users further.

6.2 Participants for Heuristics test

Test participant 1	
Name of test participant (will be anonymized)	BM
Which test items are linked test participant (e.g. questionnaires)	Appendix 6.1 Usability Report: A heuristic evaluation of the authoring tool
Gender	F
Age	38

Occupation	Research assistant
If teacher A) Which subjects?	Usability and User experience
Where was the test conducted (e.g. remotely or on site with observations)?	On site (A. C. Meyers Vænge 15, Copenhagen)

Test participant 2	
Name of test participant (will be anonymized)	LET
Which test items are linked test participant (e.g. questionnaires)	Appendix 6.1 Usability Report: A heuristic evaluation of the authoring tool
Gender	F
Age	26
Occupation	Research assistant
If teacher A) Which subjects?	Usability and User experience
Where was the test conducted (e.g. remotely or on site with observations)?	On site (A. C. Meyers Vænge 15, Copenhagen)

6.3 Participants for Visualization Test

Test participant 1	
Name of test participant (will be anonymized)	TT

Which test items are linked test participant (e.g. questionnaires)	Open and focus discussions with respect to general visualizations as described in "D2.3 Visualization strategies for course progress reports" pages 20, 24, 25, 26, 27, 28 Note taking
Gender	F
Age	39
Occupation	Teacher
If teacher B) Which subjects?	Physics, Sciences
If teacher C) Which age group	10-12
Where was the test conducted (e.g. remotely or on site with observations)?	On site

Test participant 2	
Name of test participant (will be anonymized)	GM
Which test items are linked test participant (e.g. questionnaires)	Open and focus discussions with respect to general visualizations as described in “D2.3 Visualization strategies for course progress reports” pages 20, 24, 25, 26, 27, 28 Note taking
Gender	M
Age	45
Occupation	Teacher
If teacher B) Which subjects?	Physics, Maths, Informatics
If teacher C) Which age group	10-18
Where was the test conducted (e.g. remotely or on site with observations)?	On site

Test participant 3	
Name of test participant (will be anonymized)	IA
Which test items are linked test participant (e.g. questionnaires)	Open and focus discussions with respect to general visualizations as described in "D2.3 Visualization strategies for course progress reports" pages 20, 24, 25, 26, 27, 28 Note taking
Gender	M
Age	45
Occupation	Teacher
If teacher A) Which subjects?	Physics, Sciences
If teacher B) Which age group	10-12
Where was the test conducted (e.g. remotely or on site with observations)?	On site

6.4 Questionnaire for evaluation of perceived usefulness and ease of use [11]

PERCEIVED USEFULNESS		1	2	3	4	5	6	7		NA
1. Using the system in my job would enable me to accomplish tasks more quickly Comments:	unlikely	○	○	○	○	○	○	○	likely	○
2. Using the system would improve my job performance Comments:	unlikely	○	○	○	○	○	○	○	likely	○
3. Using the system in my job would increase my productity Comments:	unlikely	○	○	○	○	○	○	○	likely	○
4. Using the system would enhance my effectiveness on the job Comments:	unlikely	○	○	○	○	○	○	○	likely	○
5. Using the system would make it easier to do my job Comments:	unlikely	○	○	○	○	○	○	○	likely	○
6. I would find the system useful in my job Comments:	unlikely	○	○	○	○	○	○	○	likely	○
PERCEIVED EASE OF USE		**1**	**2**	**3**	**4**	**5**	**6**	**7**		**NA**
7. Learning to operate the system would be easy for me Comments:	unlikely	○	○	○	○	○	○	○	likely	○
8. I would find it easy to get the system to do what I want it to do Comments:	unlikely	○	○	○	○	○	○	○	likely	○
9. My interaction with the system would be clear and understandable Comments:	unlikely	○	○	○	○	○	○	○	likely	○
10. I would find the system to be flexible to interact with Comments:	unlikely	○	○	○	○	○	○	○	likely	○
11. It would be easy for me to become skillful at using the system Comments:	unlikely	○	○	○	○	○	○	○	likely	○
12. I would find the system easy to use Comments:	unlikely	○	○	○	○	○	○	○	likely	○
		1	**2**	**3**	**4**	**5**	**6**	**7**		**NA**

List the most **negative** aspect(s):

1.
2.
3.

List the most **positive** aspect(s):

1.
2.
3.

6.5 Questionnaire for for system usability testing [12]

			1	2	3	4	5	6	7		NA
1.	Overall, I am satisfied with how easy it is to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
2.	It was simple to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
3.	I can effectively complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
4.	I am able to complete my work quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
5.	I am able to efficiently complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
6.	I feel comfortable using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
7.	It was easy to learn to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
8.	I believe I became productive quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
9.	The system gives error messages that clearly tell me how to fix problems	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
10.	Whenever I make a mistake using the system, I recover easily and quickly	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
11.	The information (such as online help, on-screen messages, and other documentation) provided with this system is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
12.	It is easy to find the information I needed	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
13.	The information provided for the system is easy to understand	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
14.	The information is effective in helping me complete the tasks and scenarios	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
15.	The organization of information on the system screens is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
16.	The interface of this system is pleasant	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
17.	I like using the interface of this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
18.	This system has all the functions and capabilities I expect it to have	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
19.	Overall, I am satisfied with this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
			1	2	3	4	5	6	7		NA

List the most **negative** aspect(s):

1.
2.
3.

List the most **positive** aspect(s):

1.
2.
3.