Chapter 6

Evaluation

This chapter describes the approaches for and the key results obtained from the evaluation of WEAVE.

Goals for the Evaluation

The main questions to be answered during the evaluation are:

* Is the system usable by pupils? The evaluation will examine how good the student interface is in communicating the worked examples effectively. It will make conclusions of whether the student interface facilitates or hinders viewing of worked examples and it will indicate potential poor design decisions for this interface to be improved in the next version of the system.
* Is the system usable by teachers? The evaluation will examine the extent to which the design decisions for the teacher interface support teachers in their goal to monitor the progress of their pupils and it will indicate potential poor decisions for this interface to be improved in the next version of the system.
* Is the usage data of pupils presented in a clear and understandable way for teachers to help them to better understand the difficulties of their pupils? The evaluation will examine how well the progress information for the students is visualised on the graphs for the teacher interface and will indicate any poor design decisions for the visualisation of data to be improved in the next version of the system.
* Is WEAVE easily deployable in schools? The major motivation for this project is to create an easily deployable tool to be used in schools. Even though no installation will be needed for the application, it is possible that schools block the website due to various access policies. This evaluation will examine how easy it is to start using WEAVE in schools.

Methods of Evaluation

Four different types of evaluation have been decided upon in order to answer the questions above as thoroughly as possible.

* Usability Evaluation conducted by participants who are not teachers or students in Computing Science. The reason for this choice of participants is so that the user interface can be judged with no bias. In other words, this evaluation aims to examine how easy to use and how self-explanatory the interface is.
* Usability Evaluation with teachers in schools. They will be able to provide some feedback on how good the application is for the intended context as well as whether it has been achieved its goal to be easily deployable in schools.
* Heuristics Evaluation conducted with two experts in Django web application development- Dr Leif Azzopardi and Mr David Maxwell- authors of the award nominated book How to Tango with Django.
* Practical Evaluation by the secondary school teachers- Mr. Peter Donaldson and Mr. Craig -who will use it as a teaching technique with their students to evaluate WEAVE more thoroughly in achieving its goals, in other words- is it easy to use WEAVE in schools, is it well-accepted by pupil and is it helpful for them as teachers to better understand the problems of their pupils.

Stages of Evaluation

Usability Evaluation

The focus of this evaluation is the quality of the user’s experience at the time of interaction with the system. To ensure that the evaluation results are not affected by the contents of the worked examples but they rather provide useful insight for the design of the user interfaces, it is conducted with both participants with and without experience in Computing Science. Comparison of the results of the two groups of participants will result in higher confidence in the conclusions drawn from this type of evaluation.

Both types of Usability Evaluation- with inexperienced and experienced participants- were conducted with six people. The choice of number of people is made on the grounds of literature which suggests that five participants are enough to discover around 80% of the usability issues in a system (reference).

Usability Evaluation with Participants without Experience in Computing

The participants were encouraged to use the tool for about half an hour and then answer two sets of questions assessing the system. During the evaluation, these participants were advised to not try and understand the meaning of the worked examples they are working on and should not worry if they are not familiar with the terminology involved in these examples.

Structure of the Evaluation

The evaluation was conducted in the MAC Lab in SAWB in the University of Glasgow, at times when no other people were present in the lab. Participants were briefed by the evaluator with information about the context the application is intended for as well as the main functions of the system. They were then given an information sheet with instructions about the goals they need to achieve using the system and one of the sets of questions they need to answer after the evaluation. This information sheet is shown in Appendix X. The reason behind providing this set of questions in advance is so that participants are aware of them and they can take notes during the evaluation if they desired so.

During their interaction with the system, participants were encouraged to think aloud. The evaluator was quietly observing them, taking notes on how they were interacting with the system as well as on their “thoughts”. The evaluator paid particular attention to signs from the user interaction providing insight on any of the following categories:

* Intuitive design. Do the users understand the architecture and the navigation of the application?
* Ease of learning. How fast did participants who never used WEAVE accomplish their goals?
* Errors and recoverability. What errors did the participants make and how quickly did they recover from them?
* Satisfaction with the system. How satisfied with the user interface are the participants?

After the participants completed the goals for the evaluation, they answered the questions in the information sheet given to them at the start of the evaluation. They also rated the system using the well-recognised System Usability Scale (SUS). Literature suggests that SUS is a quick and reliable usability assessment method (reference).

Results

Discussion on the Results

Usability Evaluation with School Teachers in Computing Science

Heuristics Evaluation

The Heuristics Evaluation of this project was conducted with two experts in Django application development from the Computing Science department in the University of Glasgow- Dr. Leif Azzopardi and Mr. David Maxwell. The main purpose of this evaluation was to have the interface compared against well-established usability principles and to identify any potential issues with the usability of the system. These principles are listed below, together with a short explanation for each.

* Visibility of the system status. This principle evaluates whether sufficient feedback for the system’s status has been provided on interaction with the system.
* Match between system and real world. This principle assesses the suitability of the terminology and the icons used in the system.
* User control and freedom. This principle evaluates the extent to which the user feels restricted by the system to achieve their goals or to navigate through the website.
* Consistency and Standards. This principle assesses how consistent the terminology and the look and feel of the system are.
* Error prevention. This principle evaluates the techniques used to reduce the places in the system where the user could enter an invalid input.
* Flexibility and efficiency of use. This principle provides an insight on how self-explanatory the interface is and whether any features for improving the efficient use of the system have been provided.
* Aesthetic and minimalist design. This principle evaluates the simplicity of the design while still being able to provide all the functionality required.
* Help users recover from errors. This principle assesses how informative the error messages provided in the system are.
* Help. This principle evaluates the quality and sufficiency of the help provided in the system.
* Navigation. This principle assesses the means of navigation across pages on the website.
* Structure of information. This principle assessed how well and logically structured different parts of the interfaces are.

The participants in this evaluation were given the same information sheet and instructions used for the Usability Evaluation with the inexperienced in Computing Science participants. They interacted with the system for about half an hour. In the end they were given an additional set of questions related to the heuristics listed above. This set of questions can be found in Appendix X.

Unfortunately, due to limited time dedicated for this evaluation, one of the participants- Dr. Leif Azzopardi, was not able to answer these questions one by one. However, he expressed his general opinion about different aspects of the user interface and suggested some areas for possible improvements. His insights are discussed in a greater detail below.

@article{Virzi:1992:RTP:141691.141700,

author = {Virzi, Robert A.},

title = {Refining the Test Phase of Usability Evaluation: How Many Subjects is Enough?},

journal = {Hum. Factors},

issue\_date = {Aug. 1992},

volume = {34},

number = {4},

month = aug,

year = {1992},

issn = {0018-7208},

pages = {457--468},

numpages = {12},

url = {http://dl.acm.org/citation.cfm?id=141691.141700},

acmid = {141700},

publisher = {Human Factors \& Ergonomics Society, Inc.},

address = {Santa Monica, CA, USA},

}

SUS @article{doi:10.1080/10447310802205776,

author = {Bangor, Aaron and Kortum, Philip T. and Miller, James T.},

title = {An Empirical Evaluation of the System Usability Scale},

journal = {International Journal of Human-Computer Interaction},

volume = {24},

number = {6},

pages = {574-594},

year = {2008},

doi = {10.1080/10447310802205776},

URL = {

http://dx.doi.org/10.1080/10447310802205776

},

eprint = {

http://dx.doi.org/10.1080/10447310802205776

}

}

In the testing chapter- describe what platforms it is tested on and how well it behaves on each of them.