**Evaluation**

In order to assess the usability of EMS, formal user evaluations were carried out in line with ethics approval obtained for this project. To determine the usability of EMS, participants are set a series of tasks on both the administration and main sides of the web-based application. Participants would complete two identical questionnaires as part of the system usability scale (SUS) assessment, with one of these tests being for the administration side of the system, and the other for the main side. Prior to the undertaking of the evaluations which contributed to the results presented in this chapter, a test run was conducted with a participant to gauge the suitability of the tasks being carried out by participants.

**System Usability Scale (SUS)**

SUS was used to determine the usability of ESM. This assessment is a multiple-choice questionnaire consisting of ten questions, each of which asks the user to respond to a statement with their opinion based on a five-point scale. The developer chose to adopt this test due to its ability to reliably provide accurate results on the systems usability even when using small sample sizes (<https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>), an important consideration given projects undertaken by students at the university often have small sample sizes. Additionally, it is straightforward to administer and places a minimal burden on participants to complete.

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

To obtain the SUS score, each odd numbered question has one subtracted from the user response, while each even number question has their score subtracted from five. The sum of all these scores is calculated and then multiplied by 2.5 with the result being the SUS score.

**User Testing**

**Testing Equipment and Location**

Testing was carried out on an Acer Aspire V laptop, with all testing taking place in the University of Dundee campus.

**Recruitment of Participants**

Participants were recruited by formal invitation to participate in the project. The inclusion criteria imposed on participants undertaking the study was that they must be at least 18 years old and presently a student, member of teaching staff, or involved in an academic exam setting process.

**Participant Tasks**

Prior to beginning testing, the participants eligibility was confirmed. Participants were issued with an information sheet and upon reading this document, were asked to complete a consent form. Participants could ask questions at any point during the assessments and were free to leave should they wish.

**Administration of Usability Test**

To test the administration site, users were tasked with performing a series of tasks which would typically be carried out by the local examination officer. These tasks include locating records, creating new deadlines, adding a new user, editing existing records, creating a module and forming a team. Upon completing these tasks, the participant completed a SUS questionnaire.

**Main Site Usability Test**

To test the main site, users were asked to perform several tasks involved in the creation of and the subsequent review process of the exam. Additionally, they carried out a password reset on the account they had created. Upon completing all tasks, participants completed a SUS questionnaire for the main site of the application.

As the system encompasses many different users the tasks carried out focused on only part of the review process. This decision was taken as to carry out all the tasks, a participant would have to repeatedly switch accounts, signing in and out each time. Doing this would give the user an experience of the system which would be far detached from the experience of an end user. There was a consideration to perform a group assessment, where each user would perform the tasks of a specific user role in the system. However, this was decided against due to the likely difficulties which would arise trying to arrange multiple participants to meet for user testing at the same time.

**Test Study**

As mentioned previously, a test study was carried out to assess the suitability of the tasks carried out by participants. The test was performed in the same manner as the actual user evaluations detailed above but results of the SUS assessment were discarded and not used in the final evaluation. This is as some minor changes were made to the task sheet issued to future participants as a result of feedback from the test study. These changes involved some minor changes to wording and separating the task list over two pages, one for the administration tasks and the other for the main site tasks. Other feedback included that having to sign in and out of accounts to complete the exam review process detracted from the immersion of the system. Subsequently, when undertaking testing the developer ensured that a second browser window was opened and signed into the other account the user needed to access when carrying out the reviews. This meant that the user could simply switch to the other browser, perform the tasks needed on the other account, then switch back to their initial browser and continue the tasks on their first account.

**Evaluation Results**

The participants which undertook the study included current local examination staff from the School of Computing, individuals involved in exam setting for other institutions and students. The total sample size consisted of seven participants. The following tables detail a breakdown of response scores by participant.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Administration Usability | | | | | | | | | | | |
|  | **Questions** | | | | | | | | | |  |
|  | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **Usability Score** |
| Participants Response | 4 | 1 | 5 | 1 | 4 | 1 | 5 | 1 | 4 | 2 | 90 |
| 4 | 2 | 4 | 1 | 5 | 1 | 5 | 2 | 4 | 2 | 85 |
| 4 | 2 | 4 | 3 | 4 | 2 | 3 | 2 | 4 | 1 | 72.5 |
| 4 | 1 | 5 | 1 | 4 | 1 | 5 | 1 | 5 | 1 | 95 |
| 5 | 1 | 5 | 1 | 5 | 2 | 5 | 1 | 5 | 1 | 97.5 |
| 4 | 2 | 3 | 1 | 5 | 1 | 4 | 1 | 4 | 1 | 85 |
| 4 | 3 | 3 | 4 | 5 | 2 | 3 | 2 | 2 | 4 | 55 |
|  | | | | | | | | | | | **82.86** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Main Site Usability | | | | | | | | | | | |
|  | **Questions** | | | | | | | | | |  |
|  | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **Usability Score** |
| Participants Response | 4 | 3 | 3 | 4 | 5 | 2 | 3 | 2 | 4 | 4 | 60 |
| 4 | 3 | 4 | 2 | 4 | 1 | 5 | 1 | 3 | 2 | 77.5 |
| 5 | 1 | 5 | 1 | 5 | 1 | 5 | 1 | 5 | 1 | 100 |
| 4 | 1 | 4 | 1 | 5 | 1 | 5 | 2 | 5 | 1 | 92.5 |
| 3 | 3 | 4 | 2 | 4 | 2 | 3 | 2 | 4 | 2 | 67.5 |
| 4 | 1 | 4 | 1 | 5 | 2 | 4 | 1 | 4 | 2 | 85 |
| 5 | 1 | 5 | 1 | 5 | 1 | 5 | 1 | 5 | 1 | 100 |
|  | | | | | | | | | | | **83.21** |

As can be seen from the tables, the administration site received a scoring of 82.86 and the main site 83.21. This indicates that users found both sides of the EMS application to be similar in terms of usability, with the main site perceived to be marginally better. Generally, a SUS scoring in excess of 68 is considered an above average in terms of usability (<https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>), which would indicate EMS performs well in this regard. It would have been beneficial to gain SUS scores of similar systems to compare EMS performance to them. However, due to the difficulties outlined in the Background chapter in locating similar systems and the unavailability of SUS scores for comparable systems, a comparison was unable to be performed.

Additionally, after discussion with the client it was agreed that there is no requirement for EMS to be compliant with Web Content Accessibility Guidelines (WCAG). As a result, it is possible that a disabled user’s SUS scoring of the system may vary dramatically from the scores recorded.

This chapter has discussed the evaluation approach adopted in assessing EMS usability. The results have been presented and their implications analysed. The following chapter will present an appraisal of the project undertaken.