Chapter # - will be assigned by editors

USABILITY EVALUATION OF VIRTUAL LEARNING ENVIRONMENTS

A University Case Study

AUTHOR(s) – Attila Vertesi, Huseyin Dogan, Angelos Stefanidis

Abstract: Virtual Learning Environments (VLEs) are widely adopted in higher education to facilitate online methods of content delivery for the lecturers, to enable online submission for the students and to provide course management tools for the administration team. There are a variety of software solutions to choose from, modelled according to heterogeneous needs and functionalities. Despite the existence of clear organisational, pedagogical, and technological goals, the procurement and implementation of a VLE is a complex task with significant challenges. We present the case study of a university which went through the process of procuring a new VLE. We examine the usability of each VLE utilizing the System Usability Scale (SUS) and capture further feedback from the stakeholders by applying the Interactive Management (IM) methodology. The first part of the research focuses on the three VLEs remaining in contention during the final stages of the procurement process. The results of the usability evaluations are analysed, explained and compared. The second part of the study examines the selected and implemented VLE six months after its initiation. A usability test was carried out again on this VLE to examine changes since its launching. Additional feedback was collected from the stakeholders to support the fine-tuning process after the implementation. According to the evaluation of each user group, all the three VLEs performed below the average usability expectation. Generally, students evaluated the usability of the VLEs higher than the academics and administration staff. The usability scores of the students' evaluation from different courses and years show remarkable differences. The ranked and categorised feedback given by the stakeholders highlights the importance of planning, training and communication prior to and during the implementation process. Usability and learnability play important roles according to the feedback.

Keywords: Virtual Learning Environment (VLE), Learning Management System (LMS), Usability Evaluation, System Usability Scale (SUS), Interactive Management (IM)

1. INTRODUCTION

Information Technology is an essential component of educational technology in Higher Education (HE). Virtual Learning Environment (VLE) and Learning Management System (LMS) are often used as synonyms [1] to describe a complex Information Technology system which integrates course management tools for course administrators, online accessibility of learning materials, and assignments. It also provides a communication and collaboration platform for the students and lecturers [2]. The quality and usability of a VLE are key features of a successful system, as they influence user satisfaction and acceptance [3]. Usability is the extent to which users can use a product or service to achieve specified goals efficiently and effectively while promoting feelings of satisfaction in a given context of use [4]. There are two aspects of usability in educational technology, namely technical usability and pedagogical usability [5]. Technical usability refers to Human-Computer Interaction (HCI), while pedagogical usability is associated with supporting the learning process. Perceived usefulness and perceived ease of use are also great influential factors in the acceptability of new technology.[6]

As part of this chapter, we present the case of a particular university which went through the process of procuring a new VLE through a tender process. We examine the migration of all of the pedagogical and administrative learning function from the old to the new system, and carefully consider the adoption of the new system by its different stakeholders. In doing so, we examine the usability of the VLEs and capture feedback from the stakeholders.

The university discussed in this research has various multiple VLEs have been in operation for more than 12 years, currently used by over 20,000 students and 2,000 staff. Due to EU procurement regulations, the university was required to go out to tender for a new VLE at the end of the contract with the current VLE supplier. In total, 250 students and staff, representing ten departments from across the university participated in the selection of the new VLE. More than 50 members of staff worked on the procuration and implementation of the new VLE for eight months before it was introduced in September 2017. During the initial phase of the implementation, in the first six months, 40% of the students were transferred to the new system. The university moved towards the full rollout a year later, in September 2018. Some of the features of the new VLE system include the provision of a personalised learning experience supported by learning analytics capabilities, integrated social media, chat, video features, and game-based learning, predominantly aimed at supporting students and their learning. As separate user group underpinned by a different set of requirements, the academic and administrative staff interacting with the system benefit from the customisable

course development, programme management, user account management, training, and end-user help desk support.

2. LITERATURE REVIEW

There is a wide range of usability evaluation methods. System Usability Scale (SUS) [7] is one of the most accepted and popular tools for measuring user satisfaction. SUS was utilised to carry out a general quantitative usability evaluation. The SUS scores from different user groups were analysed and compared. More detailed, factor analysis was applied where the low values of the usability scores required it.

While SUS gives a reliable and comparable quantitative result (SUS Score), the qualitative element of the research comprised the utilisation of an approach called Interactive Management (IM) [8], which supports the better understanding of the dynamics of the implementation process. IM was applied to facilitate effective group communication [20] to receive detailed feedback about the usability and the implementation of the new VLE. These methods are discussed in detail in the Research Methods section.

A growing number of studies examine the usability of the VLE by utilizing SUS as a methodology. In 2006, a web-based e-learning platform called SPIRAL was developed and evaluated [9] at University Claude Bernard Lyon 1. Although the SUS ratings have not been published, 72% of the professors found the system usable, according to the paper.

Three different e-learning platforms were measured using SUS by Ayad and Rigas [10]. User performance, learning effectiveness and satisfaction were examined to explore the usability aspects of the system. The three platforms were Virtual Classroom, Game-based and Storytelling. The SUS scores for the three platforms were 75.3, 73.4 and 64.5 respectively. The Storytelling scored a little behind the other two. An SUS score above a 68 would be considered above average and anything below 68 is below average.

An interesting comparative research article was published [11] regarding the usability enhancement of the Moodle LMS. The study examined the performance of the system in remote collaboration. The SUS score of the Moodle system in these features initially was 46.75, which indicates serious usability problems. Using a different collaborative tool called Drag&Share within Moodle, the usability of the LMS enhanced dramatically. The SUS score increased significantly, up to 89.5 after the implementation of Drag&Share, which indicates very good usability in the remote collaboration feature.

There is a very rare longitudinal study about a simulation-based learning system [12], that measured the perceived usability of the students after the first semester and after the second semester. Initially, the SUS score was 58.1, suggesting that the system needed improvement. Based on the collected data, the system had been modified, and after the second semester, the score rose to 65.9. Following another development for teachers, they evaluated the new module to 74.45, showing their satisfaction. This research also highlights the perceived usability of different user groups (e.g., teachers and students) may vary.

The above-mentioned divergence between the perceived usability of students and teachers is discussed by Emelyanova and Voronina [13]. The various aspects of the VLE and the difference between the perception of the usability should be considered when making a decision about the improvement of the system.

A comprehensive usability study was conducted in nine European secondary schools, all using UNITE e-learning platform, with the participation of 23 teachers and 47 students [14]. Teachers evaluated the system at 53.15 and students gave 59.36 on average using the SUS questionnaire. The difference between the perception of the usability is also noticeable in this study. However, in this case, the students scored the system higher than the teachers.

A new scale was developed by Onacan and Erturk [15] based on the SUS [7], which has been tested and validated in the HE environment for two years. The Scale for Usability of Learning Management System (SULMS) is a 26-item, Linkert-type questionnaire, which identifies five dimensions: learnability, efficiency, memorability, errors and satisfaction. In addition to SUS, SULMS tries to identify the association between the five dimensions and specific VLE-related attributes.

3. RESEARCH METHODS

3.1 System Usability Scale (SUS)

SUS can provide a simple numeric result of the perceived usability of a system from different perspectives of the diverse users and user groups [7]. The SUS scores of various systems, or the same system at different development stages, can be compared. It is easy to interpret and communicate the explicit results to the stakeholders. The evaluation is reliable even with a small sample size of 12 [16]. The survey is simple, short and there is no licence fee. These

features make SUS a perfect tool for quantitative research on the usability of VLEs. SUS is a five-point Linkert-type scale commonly applied in research which uses questionnaires. SUS includes 10 general statements regarding the user's subjective opinion and feeling of the system. The participants ranked the statements between 1 and 5 based on how much they agree or disagree with it [7]. Usability evaluation is linked to user satisfaction, enjoyment, effectiveness, efficiency.

The original statements [7] were used in the preliminary evaluation for the three VLEs (Table 2.1) and a slightly modified version was utilised during and after the implementation of the successor VLE. The first statement of the survey was rephrased from conditional tense to indicative form as the users had no option to use other VLE for these specific tasks. For the preliminary evaluation of the three different VLEs, the original phrase was used for the first statement: "I think that I would like to use this VLE frequently.". The word "system" was altered to "VLE" in every survey referring to the current system.

Table 2.1 The original statements for SUS [7]

	The System Usability Scale Standard Version		**Strongly disagree**				**Strongly agree**
			1	**2**	**3**	**4**	**5**
1	I think that I would like to use this system.		○	○	○	○	○
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 the 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.		○	○	○	○	○

Prior to completing the surveys for the preliminary evaluation of the three VLEs, the specific user groups had to perform specific, VLE-related tasks (Table 2.2) based on the most common activities they need to accomplish using the VLE. For the evaluation of the new VLE, as the system had already been used by the participants, they did not complete predefined tasks. Instead, a list of the description of various tasks offered to the participants to indicate which task had been completed by the specific user. According to Boyd et. al.

research [25] on the memory effect and recall bias, this arrangement does not influence the result of the usability evaluation.

Table 2.2 Task Lists

Student Task List	**Administrator Task List**
1 Access a unit area within the VLE	1 Navigate to a unit area
2 Review unit announcements for any notices	2 Take three word documents and make them available to students
3 View on line unit material available within the unit	3 Make a document unavailable to students
4 Open word documents made available	4 Create a link to an external website and make it available to students
5 View embedded/linked video content	5 Post an announcement to students enrolled on the unit
6 View the unit discussion topic and post an introductory message	6 Send an email to the students enrolled on the unit
7 View the unit blog and post an introductory post	7 Create a group of students for the unit
8 View the unit wiki and post an introductory page	8 View student grades and assessments
9 Complete the sample unit test	9 Access an individual Turnitin submission, view grade and feedback.
10 Submit an assignment via Turnitin	10 Add a grade for a non-Turnitin student assessment
11 View your grades	11 Add grades for all students on a non-Turnitin student assessment
12 View any notifications	12 Use the grading functionality to create a calculation which sums the Turnitin and non-Turnitin

Academics and Learning Technologists Task List
1 Take three word documents and make them available to students.
2 Make some text and an image available to students.
3 Create a link to an external website and make it available to students.
4 Make a YouTube video available to students.
5 Edit one of the items created in steps 1-4.
6 Re-organise the items previously created.
7 Make one of the items created in steps 1-4 unavailable to students.
8 Post an announcement to students.
9 Send an email to the students enrolled on the unit.
10 Create a group of students for the unit.
11 Create a discussion topic and post an introductory message.
12 Create a blog and post an introductory post.
13 Create a wiki and post an introductory page.
14 Create a test containing one multiple choice question and one multiple answer question.
15 View student grades and assessments.

16 Access an individual Turnitin submission, add a grade and feedback.
17 Add a grade for a non-Turnitin student assessment.
18 Add grades for all students on a non-Turnitin student assessment.
19 Use the grading functionality to create a calculation which sums the Turnitin and non-Turnitin assessments.

The calculation of the SUS scores of the survey was carried out by using spreadsheet software. The scores of item 1, 3, 5, 7, 9 were deducted by 1 (score - 1) and the scores of item 2, 4, 6, 8, 10 were deducted from 5 (5 - score). With this method, the positive scores given to the negative statements have been compensated by reversing the score. Now, there are ten scores ranged from 0 to 4 that gives a range of possible values from 0 to 40 in total. To extend it to a 0-100 scale, the scores were multiplied by 2.5 which gives the final SUS score of the VLE.

Experiments show [17] [18] that the average SUS score of more than 3000 different products is around 70. Specifically, for web pages and software with a web interface, this mean score is 68 which is used as a benchmark in this research.

3.2 Interactive Management (IM)

Interactive Management (IM) is a methodology designed to manage complex or new organisational or technical problems associated with multiple disciplines, involving different departments [19]. IM offers methods to facilitate effective communication, promotes consensus-based decision making through idea generation, structuring and design. These methods can be used to gather the requirements, needs, demands and ideas of the stakeholders for a better understanding of the problem space [20]. During the implementation process of a new technology, e.g., VLE, it is important to capture feedback including ideas, issues, suggestions and requirements from the users. IM can be utilised to support the qualitative part of the usability research and a better understanding of the implementation process.

In this research, IM tools are utilised to obtain feedback from the users about the implementation of the new VLE system. IM involves three phases: Planning, Workshop and Follow-up [21]. During the workshop, Trigger Questions, Idea Writing (IW) and Nominal Group Technique (NGT) were applied. The outcome of the Workshop is a list of ranked and organised statements reflecting the implementation phase of the new VLE, addressing positive and negative usability issues.

A three-hour meeting was organised by the authors in April 2018 at the university for academics (n=4), administrators (n=8) and learning technologists (n=1). The participation was voluntary. The aim of the IM session was to collect feedback, discuss questions, problems and capture ideas in connection with the implementation and usability of the new VLE.

Idea Writing

At the beginning of the IM session, the facilitator (one of the authors) introduced the methods and the Trigger Questions for the Idea Writing (IW):
Trigger Question 1:
"What are the positive aspects of the implementation of the new VLE?"
Trigger Question 2:
"What are the negative aspects of the implementation of the new VLE?"

The participants formed two mixed groups (n=6, n=7) and without discussing the question, every participant, focusing on Trigger Question 1, wrote one positive aspect of the implementation of the new VLE on his/her paper then passed the A4 sheet to the next member (on the right) of the group in the circle. After reading the previously listed statements on the new A4 sheet received from the other participant (from the left), members wrote another positive statement and circulated the A4 sheets until the original sheets arrived back. The same procedure was followed with the Trigger Question 2.

Nominal Group Technique

Following the Idea Writing phase, the members of the two groups, discussed, clarified, and edited the positive and negative statements for the preliminary ranking in each group. Each participant selected the five most important statements from the whole list and ranked them by associating numbers from one to five for each statement, five being the most important. Single Transferable Vote technique was utilised to minimise discarded votes during the ranking process.

By the end of the IM session, each group produced a list of statements to each trigger questions. The statements were discussed, clarified, and ranked. The results were photographed and transcribed. The categorising and structuring of the statements have not been accomplished due to the limited time. The results are satisfactory for providing meaningful feedback.

4. USABILITY EVALUATION AND COMPARISON

In this section, the System Usability Scale (SUS) scores are depicted in bar charts to support comparisons, analysis and the interpretation of the results. SUS scores are calculated based on the data collected from online and paper-based questionnaires. The mean SUS scores of the different user groups are compared and discussed. Further analysis was carried out where the results required it.

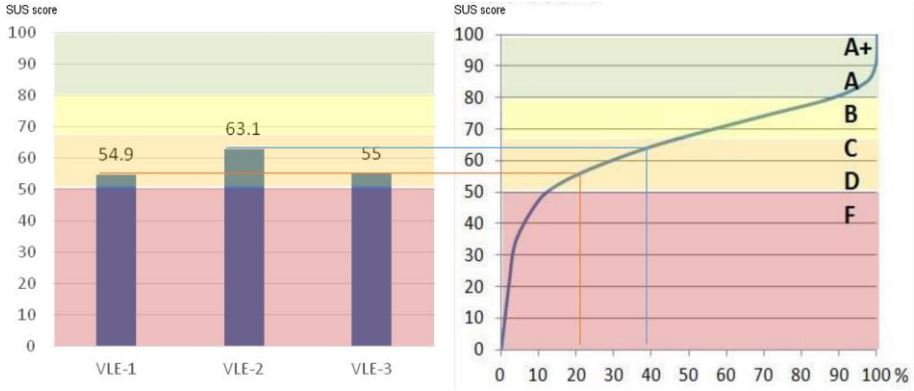
4.1 Preliminary Evaluation

The preliminary usability evaluation was carried out to support the selection process to single out one VLE from the three VLEs (VLE 1, VLE 2, VLE 3) remaining in contention during the final stages of the procurement process. The VLEs have been rated by different user groups including learning technologists (n=5), academic staff (n=32), administrator team (n=4) and students (n=40, postgraduate=13, undergraduate=22, research=5) resulting a total number of 81 SUS scores. The tasks for the usability evaluation were constructed to be in line with the role of the different user groups. The same tasks were carried out by the same groups on each of the three different VLEs (VLE 1, VLE 2, VLE 3). After the tasks were completed, participants filled in the Usability Questionnaire based on their experiences. The Usability Questionnaire includes the standard SUS questions tailored to the VLEs (Table 2.1). The same questionnaire was filled by every evaluator.

The Preliminary Results

The individual SUS scores were calculated according to the SUS methodology (Brooke 1996).

The total mean scores of the VLEs, including every user groups, are the follows: **VLE 1: 54.9 VLE 2: 63.1 VLE 3: 55.0**



3.1. Figure SUS scores (mean) of the VLEs and the percentile equivalent

The interpretation of the SUS scores

3.1. Figure 3.1 shows the results with a graphical aid for interpretation and adjective rating.

The background colours of the chart (Figure 3.1) and the letter-grades (from A+ to F) are related to the well-accepted adjective scale (Figure 3.1) based on the benchmarks set up by Bangor et. al. [18]. A SUS score higher than 80 suggests a very good, highly usable system (A+, A), between 68 and 80 still refers to a good system with space for improvement (B), between 51 and 68 means "Fair"or "OK", the system or product is still usable but should be improved (C, D), and below 51 is poor (F), below 36 is unusable.

Figure 3.1 aids to convert SUS score to percentile rank [18] by normalizing the scores based on the distribution of all scores measured in different products and systems by different users. A SUS score of 68 would be equivalent to 50% which means that the averages SUS score of all the products measured with SUS method is around 68.

VLE 1 and VLE 3 scored similarly (54.9, 55) and VLE 2 has higher score (63.1) but all the three VLEs are within the 51 - 68 range which suggests that there are no major issues with the usability but there is space for improvement.

The evaluation by user groups

The following table (Table 5.1) shows the numeric results of the evaluations of the different user groups: students (n=40), academics (n=32), learning technologists (n=5) and administrator team (n=4).

Table 4.1 The usability scores of different user groups

SUS Scores	Students (n=40)	Academics (n=32)	Administrator Team (n=4)	Learning Tech. (n=5)
VLE 1	**53.4**	**55.3**	**63.8**	**57**

VLE 2	**68.4**	**57.2**	**56.3**	**64**
VLE 3	**62.9**	**49.9**	**20.6**	**52**

The charts in Figure 4.2 offers a more detailed insight by displaying the VLE scores of each user group.

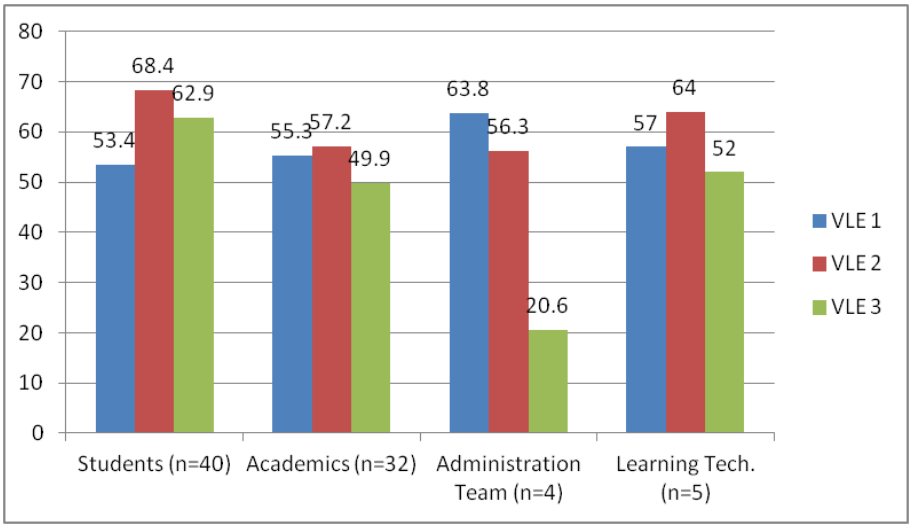


Figure 4.2 Usability Scores of different User Groups Chart

The students did not find VLE 1 as usable as VLE 2 and VLE 3. The difference between the SUS score of VLE 1 and VLE 2 is 15 which clearly shows students' preference (VLE 1). While the Academics produced more balanced SUS scores, the Administrator Team's evaluation demonstrates the highest deviation. Examination of the data reveals that there are only four members of the administrator team participated in this evaluation, and the individual scores (47.5, 2.5, 27.5, 5) show high inconsistency in the case of VLE 3. A number of studies proved that the sample size below five cannot give reliable result in usability testing, although, they still can unveil 80% of the system's usability problems [22]. Therefore, the result of the Administrator Team should still be considered as the SUS scores are all below 50 alerting to significant usability issues in their field using VLE 3. The averages of the SUS scores given by Learning Technologists are more coherent. VLE 2 performed the best by most of the user groups and the majority of the participants (n=77) except the Administration Team (n=4).

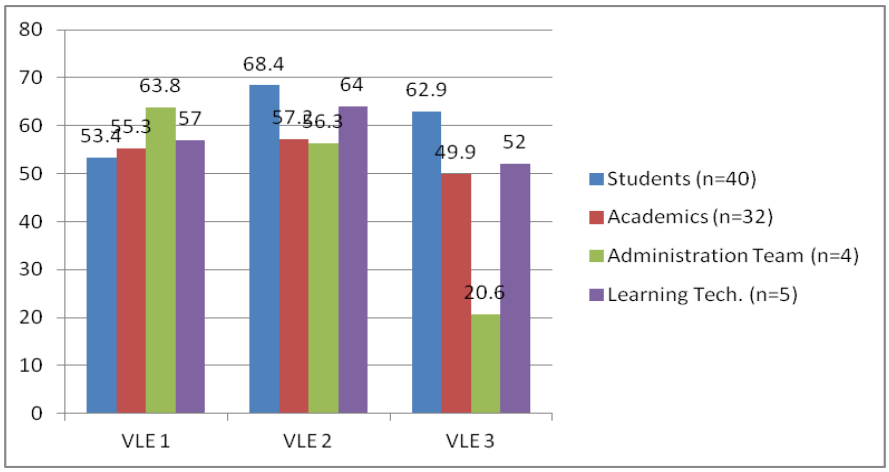


Figure 4.3 VLE Chart - User Groups

Displaying the group's SUS scores grouped by the VLEs (Figure 4.3) highlights the difference between VLE 1 and VLE 3. Even the average SUS scores were almost equal (VLE 1=54.9, VLE 3=55.0) the standard deviation shows significant differences (VLE 1=4.5, VLE 3=18). VLE 3 carries internal tension: the students graded to 62.9, 20.6 was given by the administrator team. In contrast, VLE 1 was valued at 63.8 by the administrator team and at 53.4 by the students which is the lowest score from those the student gave to the three VLEs. VLE 1 is still well balanced in the mean SUS scores of the groups. VLE 2 performed the best according to the students, academics and learning technologists and in mean score. Only the administrator team ranked VLE 2 below VLE 1.

The reliability of the evaluation

Cronbach's Alpha measures reliability by calculating the internal consistency of the data [23]. The calculation gives a result between 0 and 1. The closer the α to 1, the better. SUS performs well $\alpha > 0.9$ which means the test measures what it should be, the usability. Generally, $\alpha > 0.7$ is accepted as a reliable test consistency. The standard deviation and reliability calculations have been carried out in all test published in this paper to verify the internal consistency. The following table (Table 5.2) shows the results.

Table 4.2 Reliability of the Preliminary Surveys

	SUS Score	Cronbach's Alpha	Standard Deviation
VLE 1	54.9	0.91	22.34
VLE 2	63.1	0.92	23.34
VLE 3	55.0	0.94	25.27

4.2 The subsequent evaluation of the new VLE

The qualitative usability evaluation was conducted on the new VLE by utilising SUS methodology the same way as in the case of the preliminary evaluation. The total number of participants is n=182 including students (n=137), academics (n=23), learning technologists (n=3) and administrator team (n=19). Printed (paper) and online questionnaires were offered. N=13 SUS evaluations arrived on paper evaluated by learning technologists (n=1), academics (n=4) and administrators (n=8). Students did not participate in this session. The online questionnaire was submitted by 169 users including learning technologists (n=2), students (n=137), academics (n=19) and administrators (n=11).

In this case, the participants were not asked to complete any task prior to the questionnaire but a list of features was attached to the paper questionnaires enabling the participant to indicate which tasks have been carried out by them. The user evaluation was based on the general experience gained during the first phase of the implementation (from September 2017 to April 2018) of the new VLE by using the features needed for the different user groups. This approach does not influence the outcome of the evaluation [25].

The SUS questions were intended to be the same as the preliminary questions based on the original SUS questions [7] with a slight change in the wording. Unfortunately, a small error slipped into the online student questionnaire. Question 5 (Table 2.2) was repeated twice and as a result, the last question (Q 10) was left out. This small discrepancy does not affect the result significantly as the structure of the SUS questions and the methodology make the evaluation robust and resilient to small errors and changes [24]. The standard error is within 0.25 regarding the final SUS score. The accuracy is higher than 99.5 %.

The result

The overall SUS score of the adopted VLE is 58.6 out of 100 measured six months after the first phase of the implementation in April 2018. This is the result of the evaluation of n=182 users including students (n=137), administrators (n=19), academics (n=23) and learning technologists (n=3).

The final score does not differ significantly from the SUS scores in the preliminary evaluation. It is still in the range of 51-68 being below the average usability expectation (68) but still envisions a usable system with a scope for improvement.

Comparisons of user groups' evaluations

A more differentiated picture can be seen by examining and comparing the evaluation of the different user groups. The largest number of users participated in this evaluation are the students (n=137) scored 61.1 opposed to all members of staff (n=45) 49.4. As a result, students' SUS score weighted more in the overall score and scored 58.6 for the total average. If the two user groups formed by the students and the staff are weighted equally, the mean SUS score is 55.6, lower than the average score 58.6 calculated with all users as one group. The following chart (Figure 5.5) displays the SUS score in respect to the two main user groups, the group average and the total average.

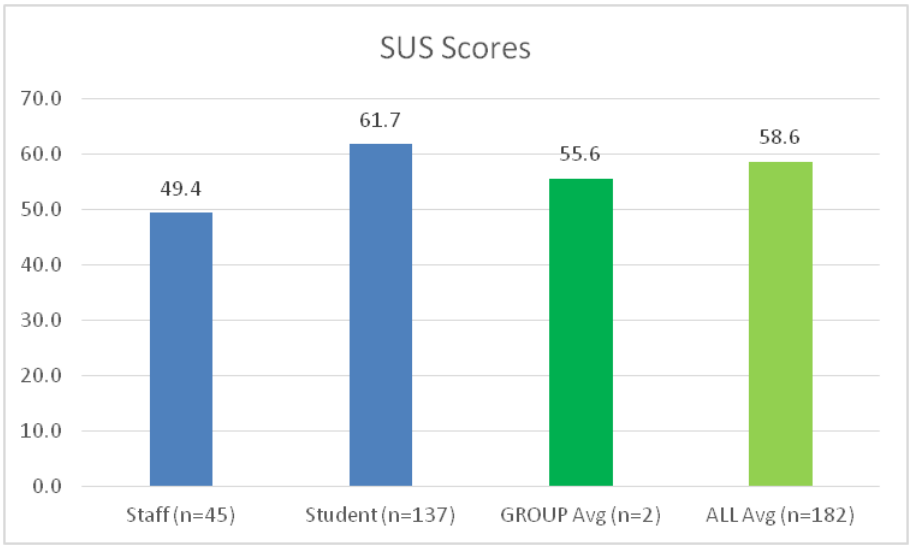


Figure 4.4 SUS Scores of Students and Staff

Students Evaluations

Starting the analysis with the largest user group, the students (n=137), it is interesting to see the comparison of the SUS scores of the different sub-groups within the students.

Student Groups by Years

Undergraduate (n=127) and postgraduate (n=10) students filled in the online form.

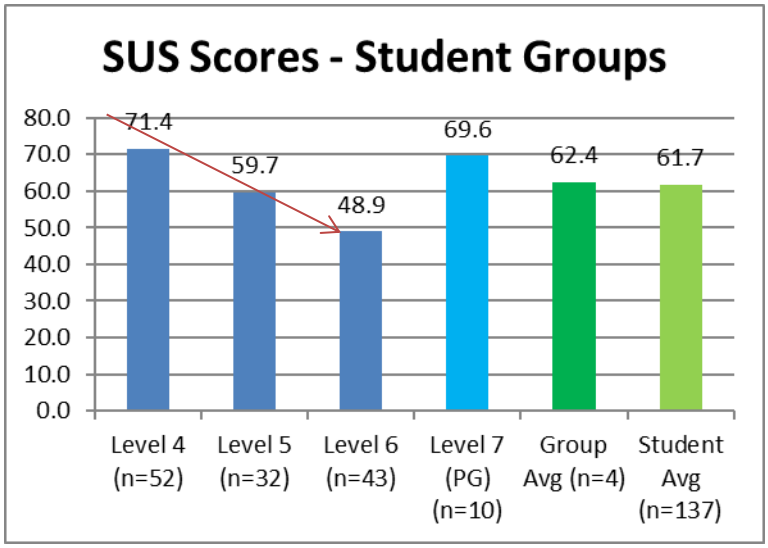


Figure 4.5 SUS Scores of Student Groups by Levels

There is a falling trend can be seen in the graph (Figure 5.6) by the undergraduate student groups from 71.4 (Level 4, Year 1) through 59.7 (Level 5, Year 2) to 48.9 (Level 6, Year 3). The Year 1 (Level 4) students evaluated the new VLE slightly above the average expectation. They seem to be more satisfied with the new system, unlike the Year 3 (Level 6) students who expressed higher expectations. The postgraduate students (level 7), however, gave 69.6 for usability (Figure 5.6) which is above the generally accepted average (68) for SUS scores.

Student Groups by Frameworks/Courses

The results of six different groups of students can be seen in Figure 5.7. The groups were formed based on curriculum areas and courses. The students participated from every year and level in each group. The largest group is the nursing students (n=66). Their average SUS score is 60.3 which is very close to the average score of the six groups (60.2) indicated by the red line in Figure 5.7. The difference between the lowest (43.9) and highest SUS score (74.7) is more than 30 (30.8).

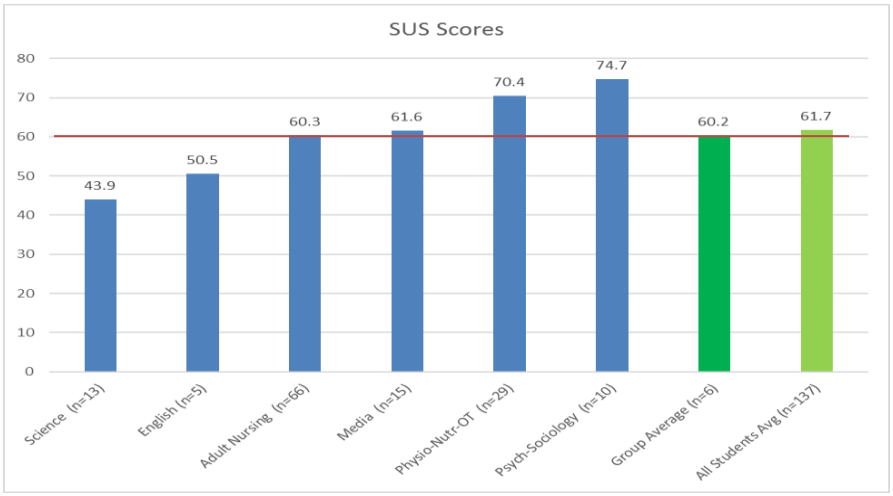


Figure 4.6 SUS Scores of Student Groups by Frameworks/Courses

Staff Evaluation

N=45 evaluation arrived from staff members either online (n=32) and on paper (n=13). The following groups are created: academics (n=23), administrators (n=20), learning technologists (n=3)

Figure 5.8 shows the results graphically. It is noticeable that academics gave very low usability score (37.8) to the new VLE since the evaluation of administrators (59.9) and learning technologists (68.3) suggest that the VLE is closer to an average system with respect to the usability. The mean value of the groups' SUS scores is 55.4 which is acceptable but the total average falls slightly below 50 (49.4) which is the minimum usability requirement of any system.

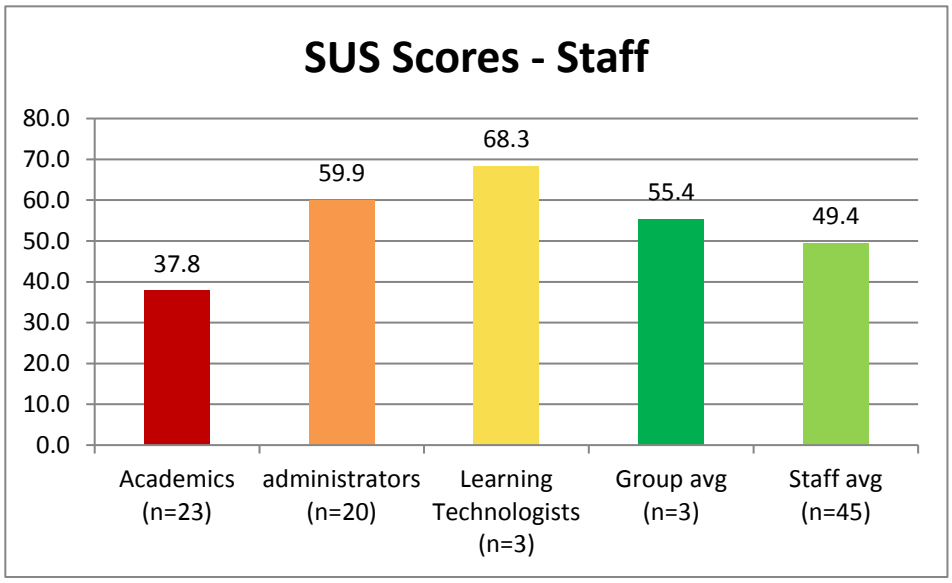


Figure 4.7 SUS Scores of Staff Groups

The result of the academics (SUS = 37.8) draws attention to some significant usability issues. For further analysis, the chart in **Error! Reference source not found.** shows the individual scores in the academics group (n=23). Blue bars (n=19) shows the online result, yellow bars (n=4) relate to the paper-based questionnaire.

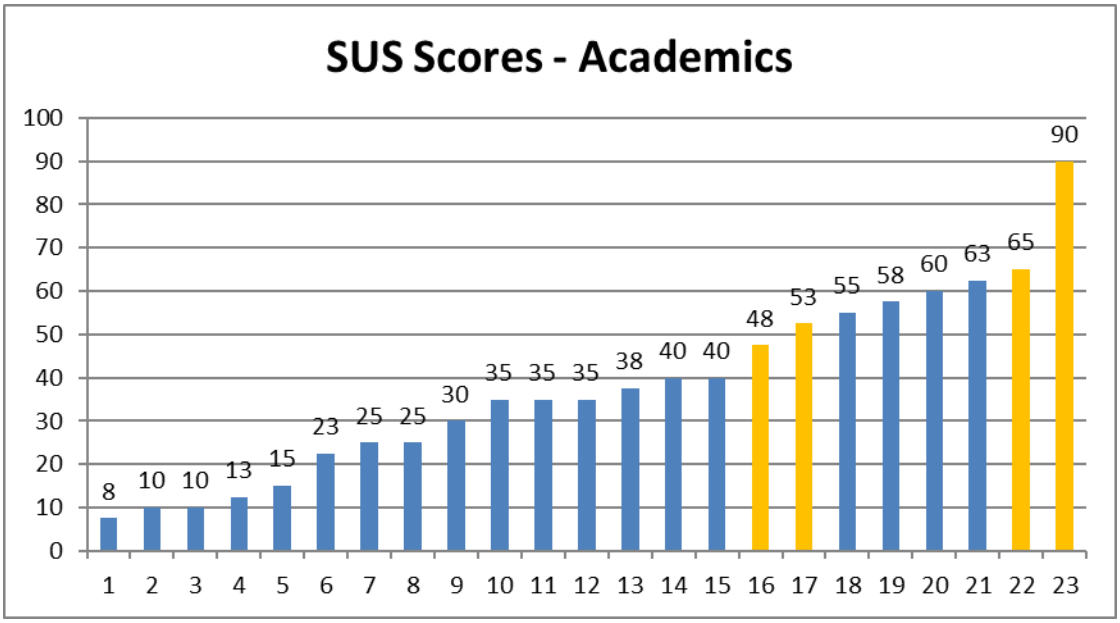


Figure 5.9 SUS Scores of Academics

Half of the group of academics (n=12) evaluated the new VLE below 38 which indicate serious usability issues. Interestingly, the paper-based results (n=4) are significantly higher (SUS avg = 64) than the online scores (SUS avg

= 32). Although, the overall standard deviation is not high (21), the range and distribution of the scores are unusual.

Factor Analysis

SUS is not a diagnostic tool, it is an evaluation method. SUS reveals but does not specify the usability problem. However, more detailed analysis can give some hints about the weak areas of the new VLE.

According to the academics (n=15) who evaluated the system lower than 41. Figure 5.10 shows the result of each factor (the scores given to each question) of the evaluations which have the total SUS score under 41. These are the first 15 scores from the left on the previous bar chart in Figure 5.9.

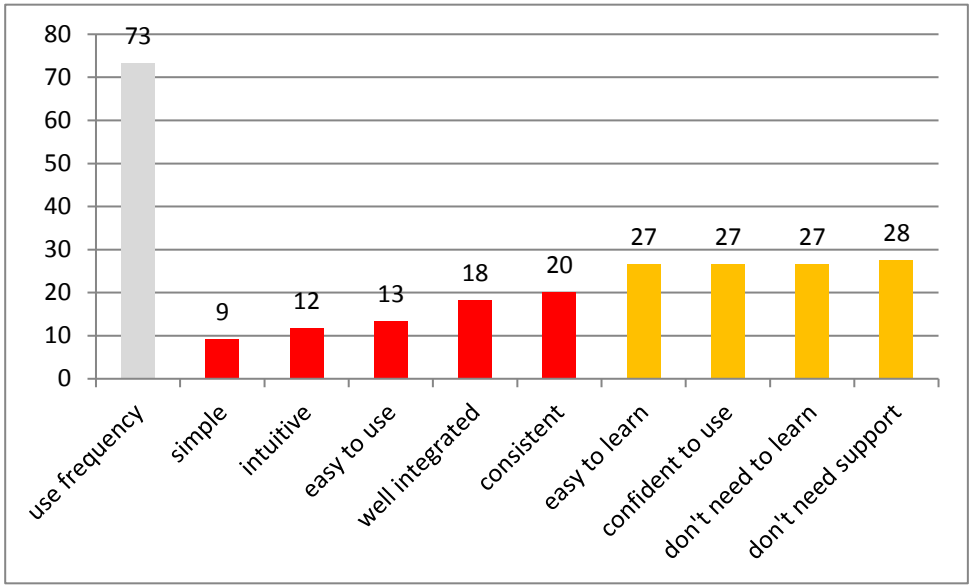


Figure 4.8 Factor Analysis of the weakest evaluations (SUS < 41)

The weakest areas are highlighted in red on the bar chart (Figure 5.9). These academics (n=15 out of 23) did not find the new VLE simple to use, intuitive, easy to use, well-integrated and consistent. The high value of the "frequency of the use" is an outlier and can be misleading as there is no alternative VLE to use at this stage. That is why it is colour coded with grey indicating the insignificance of that value.

The Reliability of the Test

The following results in the table (Table 5.3) show the mean SUS score, the reliability, and the standard deviation of the scores. Cronbach's Alpha relates to the internal consistency of the answers. The range of the reliability

score can be from 0 to 1. The closer to 1, the more reliable the result. Although the interpretation of the reliability depends on the system, usually, above 0.7 is acceptable, 0.8 is good, 0.9 refers to a highly reliable set of results. The results in the table (Table 5.3) suggest very good internal consistency. The standard deviation shows consistent result as well.

Table 5.3 Reliability of the new VLE Surveys

Reliability	**SUS Score**	**Cronbach's Alpha**	**Standard Deviation**
Students	**61.66**	**0.89**	**20.80**
Staff	**49.43**	**0.88**	**21.62**
All (Students + Staff)	**58.61**	**0.90**	**21.65**

5. INTERACTIVE MANAGEMENT SESSION

5.1 Idea Writing and Rating

By the end of the IM session, four lists of ranked statements were produced by the two groups in response to the two trigger questions. The two positive and two negative lists were merged into one positive and one negative list. The merged list can be seen in Table 5.4 of the positive and Table 5.5 of the negative statements. The wording follows the original transcript. The scores in the tables are the sum of the individual scores given by the participants. They show the importance of the statements according to the participants.

Table 5.1 Ranked List of Positive Statements

	All Positives - merged	**Score**
1	Clean and Fresh, works good, better user interface	23
2	On-demand help from Learning Tech, contact directly	19
3	Able to contact trainers	18
4	Functionalities for staff/students	18
5	Programme Support help area now a lot cleaner	17
6	Access to Sandbox to mess around without worrying about breaking the system.	16
7	Learning technologists were very helpful above expectations	15
8	Allows students to hand in late submissions in same area, lateness is clearly marked	14
9	Impersonating a student doesn't log you out	11
10	Training organised and run in plenty of time	10
11	Advantage in piloting is confidence in year 2	9

12	Not having to login	9
13	Lots of training available/given	8
14	Clear	5
15	Help section divided for academics/professional support	5
16	Open more than one screen at a time without needing to logout	5
17	Quick, intuitive so means mistakes by others easy to correct	5
18	Drag and drop files	2
19	Similar concepts to the previous VLE in terms of content structure	2

Table 5.2 Ranked List of Negative Statements

	All Negatives - merged	**Score**
1	Trainers had limited time to learn themselves	34
2	Implementation rushed meaning having to deal with issues that now arise	23
3	No LT support	15
4	Current VLE and new VLE not always linked up	13
5	Systems not talking to each other as well as advertised	12
6	Who was consulted regarding large file submission	11
7	Lack of info prior to roll out	10
8	Training too general	8
9	Learning Tech team restructured during launch	8
10	Training for turnitin not available at the time of implementation - given too early - academics need to be reminded to read help pages	8
11	A lot of things shown were not useful in terms of usability for teaching	6
12	Anticipating members of staff to be able to sort IT related issues due to incompatibility of videos, documents and live streaming apps	6
13	No personal training for unique faculty needs	5
14	More communication required about Implementation	5
15	No template for structure of unit	5
16	LT consultation at same time as everything new meant lack of support	5
17	Too many ways of accessing the same thing	4
18	Sandbox can't simulate everything	4
19	Grader app not supported	3
20	Too many courses in initial rollout	3
21	Help and guidance very lengthy and difficult to follow	3
22	Learning 'how to' at same time as LTs who often don't know how to do things	3

23	Did not have choice	2
24	Software lacks consistency in interface	2
25	Interface too 'flat' - how do you know where you are?	2
26	No confidence in software	2
27	Lack of updates when a process changes	2
28	Signposting students to new VLE - need much more	2
29	Student Support and Academics are not in the same training	1

Table 5.4 and Table 5.5 show the prioritised list with the total scores summed from the individual ranking scores.

5.2 Categories

The positive and negative statements are grouped into categories based on similarities which makes the problem domain clearer and easier to recognise structure and pattern.

The following groups (Table 5.6; Table 5.7) are created form the lists in Table 5.4 and Table 5.5. The order of the statements follows the scores in ranking. The list starts with the most important statements. Some statements are listed in more than one category if it was required.

Table 5.3 Positive Statements Grouped by Categories

Positive Statements Grouped by Categories
Usability
Clean and Fresh, works good, better user interface
Functionalities for staff/students
Programme Support help area now a lot cleaner
Allows students to hand in late submissions in same area, lateness is clearly marked
Impersonating a student doesn't log you out
Not having to login
Open more than one screen at a time without needing to logout
Quick, intuitive so means mistakes by others easy to correct
Similar concepts to the previous VLE in terms of content structure
Drag and drop files
Learnability
Access to Sandbox to mess around without worrying about breaking the system.
Training organise and run in plenty of time

Advantage in piloting is confidence in year 2
Help section divided for academics/professional support
Support
Able to contact trainers
On-demand help from Learning Tech, contact directly
Programme Support help area now a lot cleaner
Learning technologists were very helpful (HSS+FMC but not FM) above expectations
Training organised and run in plenty of time
Lots of training available/given
Help section divided for academics/professional support

Table 5.4 Negative Statements Grouped by Categories

Negative Statements Grouped by Categories
Time (time-pressure)
Trainers had limited time to learn themselves
Implementation rushed meaning having to deal with issues that now arise
Training for turnitin not available at the time of implementation - given too early - academics need to be reminded to read help pages
LT consultation at same time as everything new meant lack of support
Structural and Organisation
Learning Tech team restructured during launch
LT consultation at same time as everything new meant lack of support
Too many courses in initial rollout
Who was consulted regarding large file submission
Support
No LT support
Training too general
Training for turnitin not available at the time of implementation - given too early - academics need to be reminded to read help pages
No personal training for unique faculty needs
Too many ways of accessing the same thing
Sandbox can't simulate everything
Help and guidance very lengthy and difficult to follow
Learning 'how to' at same time as LTs who often don't know how to do things
Usability

Current VLE and new VLE not always linked up
Systems not talking to each other as well as advertised
A lot of things shown were not useful in terms of usability for teaching Anticipating members of staff to be able to sort IT related issues due to incompatibility of videos, documents and live streaming apps
No template for structure of unit
Software lacks consistency in interface Who was consulted regarding large file submission
Communication
Systems not talking to each other as well as advertised
Who was consulted regarding large file submission
Lack of info prior to rolling out Training for turnitin not available at the time of implementation - given too early - academics need to be reminded to read help pages
More communication required about Implementation
Lack of updates when a process changes Who was consulted regarding large file submission

The categories refer to usability, learnability, support, and communication. The individual statements specify the area and nature of the usability issues. IM offers valuable feedback by supporting the general evaluation of the SUS with specific comments.

6. DISCUSSION

6.1 Preliminary Results

The results show that regarding the usability, there are no big differences between the three VLEs. VLE 1 and VLE 3 reach almost identical SUS scores (54.9 and 55.0) while VLE 2 received 8 points higher score (63). The average, normalised usability score generally for web-based systems is 68. All three VLEs performed under the average expectation. There are differences in the perceived usability of the different user groups. The influential factors that

could cause differences in the results are not researched in this study. The students seem to be more satisfied with the VLEs than members of staff.

6.2 The new VLE

The usability evaluation of the new VLE at this stage provided reliable and meaningful feedback. The overall SUS score (58.6) suggests a usable system in general but also indicate some usability issues in particular areas. As the implementation is in its early stage (phase 1), this score should not be considered as a final SUS score of the fully implemented and fine-tuned system. The analysis of the evaluation of the different user groups and individual users discloses more details and differences within and between the usability perception of the user groups. The VLE is a complex system with numerous features. Each user group evaluates a slightly or significantly different part of the VLE. The divergence between the SUS scores hints that (a) the system is not uniform regarding the usability (b) the expectation and perception are different. The detailed analysis of the low SUS scores (37.8) given by the academics identified five problematic areas: simplicity, intuitiveness, ease of use, integration, consistency. Students are mostly satisfied with the new VLE, although, interesting trends can be seen in the undergraduate results (Figure 5.6). Academics and administrators are not always fully satisfied.

6.3 Interactive Management Workshop

The IM workshop offered a valuable opportunity to identify, communicate and resolve some serious usability issues. The feedback captured during the workshop was useful for the team that administer the implementation.

The feedback captured in the IM session suggests some explanation for low SUS scores. There are more negative statements (n=29) in the ranked lists than positive ones (n=19). The categories refer to areas that need attention either from the usability perspective or regarding the implementation process. The high importance of providing support, offering training and maintaining good communication is well recognized by the team that manages the implementation and confirmed by the result of this study as well.

6.4 Limitation

The research has the following limitations. The different user groups were not represented in equal number. Three times more student completed the online evaluation, but no students participated in the IM session. The SUS

score comparison of the user groups gives an equal weighting to every user group.

The phrasing of the first question was modified from ‘I think that I would like to use this system frequently’ to ‘I use this VLE frequently’. The reason behind this change is that there was no choice of using other VLE for these users. The impact of this change is that the SUS scores given to the first question are relatively high compared to the average scores. It slightly raises the mean SUS score.

6.5 Future Work

There are only a few longitudinal studies that measure the usability periodically during the whole lifecycle of the product. Even fewer usability research can be found on VLEs [12]. Even if any kind of usability evaluation is involved, usually, it is limited to the development phase only. This could suggest the assumption that the users’ perceived usability does not change after a product e.g., a VLE is installed and works properly. Even if the system does not change, the users are changing.

It would be beneficial to get feedback on a regular base regarding the user acceptance and perceived usability of the new VLE. Valuable information can be collected regarding the whole educational technology system including users and developers. A simple usability evaluation such as SUS can measure the impact of any changes on the VLE. As the perceived usability depends not only the VLE but it is influenced by the user’s attitude and other social changes, it would be even more interesting to see the effect of the pedagogical and social intervention on the SUS scores.

6.6 Conclusion

The usability evaluation provided meaningful, easy to understand and comparable results to support the decision-making during and after the procurement process regarding the three VLEs. The adopted VLE has been reevaluated six months after the introduction of the new system, at the first stage of the implementation. More detailed research comprising IM methodologies offered a realistic, specific and accurate picture about the adopted VLE at the end of the first phase of the implementation. The case

study also demonstrated an example of a feasible, quantitative and qualitative usability evaluation of a VLE combining SUS and IM methodologies.

REFERENCES

[1] Paulsen, M. F. (2002). Online Education Systems: Discussion and definition of terms. *NKI distance education*, *202*.

[2] Ryan, S., Scott, B., Freeman, H., & Patel, D. (2013). *The virtual university: The internet and resource-based learning*. Routledge.

[3] Babić, S. (2012). Factors that influence academic teacher's acceptance of e-learning technology in blended learning environment. *E-learning-organizational infrastructure and tools for specific areas*, 3-18.

[4] Iso, W. (1998). 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). *The international organization for standardization*, *45*(9).

[5] Melis, E., Weber, M., & Andrès, E. (2003). Lessons for (pedagogic) usability of eLearning systems. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 281-284). Association for the Advancement of Computing in Education (AACE).

[6] Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science*, *35*(8), 982-1003.

[7] Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

[8] Broome, B. J., & Keever, D. B. (1986). Facilitating Group Communication: The Interactive Management Approach.

[9] Renaut, C., Batier, C., Flory, L., & Heyde, M. (2006). Improving web site usability for a better e-learning experience. *Current developments in technology-assisted education*, 891-896.

[10] Ayad, K., & Rigas, D. (2010). Comparing virtual classroom, game-based learning and storytelling teachings in e-learning. *International Journal of Education and Information Technologies*, *4*(1), 15-23.

[11] Marco, F. A., Penichet, V. M. R., & Gallud, J. A. (2013). Collaborative e-Learning through Drag & Share in Synchronous Shared Workspaces. *J. UCS*, *19*(7), 894-911.

[12] Luo, G. H., Liu, E. Z. F., Kuo, H. W., & Yuan, S. M. (2014). Design and implementation of a simulation-based learning system for international trade. *The International Review of Research in Open and Distributed Learning*, *15*(1).

[13] Emelyanova, N., & Voronina, E. (2014). Introducing a learning management system at a Russian university: Students' and teachers' perceptions. *The International Review of Research in Open and Distributed Learning*, *15*(1).

[14] Granić, A., & Ćukušić, M. (2011). Usability testing and expert inspections complemented by educational evaluation: A case study of an e-learning platform. *Journal of Educational Technology & Society*, *14*(2), 107-123.

[15] Onacan, M. B. K., & Erturk, A. (2016, October). Usability evaluation of learning management system in higher education institution: A scale development study. In *International Conference on Leadership, Technology, Innovation and Business Management (ICLTIBM-2016), Antalya, Turkiye*.

[16] Tullis, T. S., & Stetson, J. N. (2004, June). A comparison of questionnaires for assessing website usability. In *Usability professional association conference* (Vol. 1).

[17] Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, *24*(6), 574-594.

[18] Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114-123.

[19] Broome, B. J., & Keever, D. B. (1986). Facilitating Group Communication: The Interactive Management Approach.

[20] Dogan, H., & Henshaw, M. J. D. (2010, April). Transition from soft systems to an enterprise knowledge management architecture. In *International Conference on Contemporary Ergonomics and Human Factors* (pp. 13-15).
[21] Warfield, J. N., & Cárdenas, A. R. (1994). *A handbook of interactive management* (p. 338). Ames: Iowa State University Press.
[22] Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough?. *Human factors*, *34*(4), 457-468.
[23] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297-334.
[24] Sauro, J., & Lewis, J. R. (2011, May). When designing usability questionnaires, does it hurt to be positive?. In *CHI*(pp. 2215-2224).
[25] Boyd, K., Bond, R., Vertesi, A., Dogan, H., & Magee, J. (2019). How People Judge the Usability of a Desktop Graphic User Interface at Different Time Points: Is there Evidence for Memory Decay, Recall Bias or Temporal Bias?. *Interacting with Computers*, *31*(2), 221-230.

ACKNOWLEDGEMENTS

The research team would like to express gratitude to the Centre for Excellence of Learning (CEL) and Visual for Learning (V4L) team of the university, especially to Ms Wendy Drake and Mr Giles Ashton for the contribution of the research by organizing surveys and providing data for the research.

AUTHOR INFORMATION

Attila S. Vertesi

Institutional affiliation: Bournemouth University

Institutional address: Fern Barrow, Poole, Dorset, BH12 5BB, United Kingdom

Complete mailing address: P206 BPC North Rd, Bournemouth University, Fern Barrow, Poole, Dorset, BH12 5BB, United Kingdom

Telephone number: +44 (0) 1202 205313

Email address: avertesi@bournemouth.ac.uk

Attila Vertesi is a Lecturer at Bournemouth and Poole College (BPC) and a Honorary Lecturer at Bournemouth University (BU) in Computing and Engineering. Attila is a Course Leader of the Access to HE Computing Course at BPC. He received his MSc in Cyber Security and Human Factors from BU, graduated as a Physics Teacher (BSc Hons) at ELTE Budapest and holds Communication Engineering (BSc) and Technical Teacher (BSc) degrees.

Dr Huseyin Dogan

Institutional affiliation: Bournemouth University

Institutional address: Fern Barrow, Poole, Dorset, BH12 5BB, United Kingdom

Complete mailing address: Poole House P429, Talbot Campus, Fern Barrow, Poole, BH12 5BB
Telephone number: +44 (0) 01202 962491

Email address: hdogan@bournemouth.ac.uk

Website: https://staffprofiles.bournemouth.ac.uk/display/hdogan

Dr Huseyin Dogan is a Principal Academic in Computing at Bournemouth University (BU). Dr Dogan's research focuses on Human Factors, Assistive Technology, Digital Health and Systems Engineering. He is the Co-Founder and Co-Chair of the Human Computer Interaction (HCI) research group and previously led the postgraduate courses (2012-2015) and then the undergraduate courses (2015-2017) in Computing at BU, managing over 900 students. He is also the Co-Founder of four MSc programmes: (1) MSc Digital Health; (2) MSc Digital Health & Artificial Intelligence; (3) MSc Cyber Security & Human Factors; (4) MSc Marketing & User Experience.
Dr Dogan received his Engineering Doctorate (EngD) in Systems Engineering from Loughborough University, MSc in Human Computer Interaction with Ergonomics from University College London, and BSc in Computer Science from Queen Mary University of London.
He was also the General Co-Chair for the 30th International British Computer Society Human Computer Interaction Conference. He has over 100 publications and his research on Assistive Technologies (with Dr Paul Whittington) featured on the BBC South, BBC Radio Solent, The Ergonomist, Auto Express, Bournemouth Echo and The Sunday Times magazine.

Dr Angelos Stefanidis

Institutional affiliation: Bournemouth University

Institutional address: Fern Barrow, Poole, Dorset, BH12 5BB, United Kingdom

Complete mailing address: Poole House P313, Talbot Campus, Fern Barrow, Poole, BH12 5BB

Telephone number: +44 (0) 1202 965500

Fax number (optional):

Email address: astefanidis@bournemouth.ac.uk

Website: https://staffprofiles.bournemouth.ac.uk/display/astefanidis

Dr Angelos Stefanidis is the Associate Dean (Global Engagement) in the Faculty of Science & Technology, Bournemouth University. He received his PhD in Information Systems Education from Loughborough University. Angelos has held various academic roles over the years and has extensive teaching experience in the areas of Computing and Information Management. He also has extensive experience in the design, management and delivery of professionally accredited CPD courses to IT companies and organisations. In addition, he has held various academic management positions and worked externally with numerous universities in the UK and overseas. His current role involves providing leadership, operational development and delivery of the Bournemouth University's Global Engagement plan by supporting recruitment, international partnerships, students and staff mobility, education, research and professional practice activities on and off-campus.