Enhanced System Usability Scale for Adaptive Courses

Tansu Pancar, Matthias Holthaus, Per Bergamin
UNESCO chair on personalized and adaptive distance education
Swiss Distance University of Applied Sciences (FFHS)
Brig, Switzerland
e-mail: tansu.pancar@ffhs.ch, matthias.holthaus@ffhs.ch, per.bergamin@ffhs.ch

***Abstract* – The usage and complexity of learning management systems is increasing continuously. The System Usability Scale (SUS) is a well-known and widely used scale to measure usability of web applications. It is also used in the context of e-learning and several applications were done in popular learning management systems. This study uses a version of the (SUS) translated from English to German on adaptive courses in Moodle and compares the results with the results from a previous study on non-adaptive courses. In addition to comparing results from adaptive and non-adaptive courses, this study also tests the two-factor structure consisting of usability and learnability and suggests two additional items to be included into the learnability factor. In the evaluation of adaptive courses, the total SUS value is lower than non-adaptive courses at our university. In particular, the SUS value for the learnability factor is far lower. It also shows that some items of our German version of the SUS are not suitable for evaluating adaptive courses.**

Keywords-e-learning; usability evaluation; adaptive learning; system usability scale; learnability; moodle.

I. INTRODUCTION

Along with the advances in technology, the interaction between education and technology is increasing all over the world. E-learning platforms and applications are being used by many institutions in pure online or blended learning scenarios. Usability of these online tools is an important factor influencing the learning outcome and satisfaction of learners. The System Usability Scale (SUS) [1], with its 10 items, is one of the shortest and most frequently used scales to evaluate web applications. This study uses a version translated into German, which has been validated once before, in our previous study for non-adaptive courses [2].

This current study contains data from university students, participating in adaptive courses in a blended-learning environment in the autumn semester 2018/19. In a first step, the data was analyzed to confirm reliability and validity. In a second step, the results of this study were compared to the results from non-adaptive courses in the previous study. For this, reliability, validity, SUS scores and factor structure were examined and compared. In the original SUS, the two-factor structure contains 8 items for the usability factor and only 2 items for the learnability factor, which does not seem a good distribution. Therefore, the third step was to extend the 10-item scale with two additional questions in order to enhance the learnability factor. Results from the original 10-item scale were compared to the 12-item scale.

Summarizing: the main goal of this study was to measure the usability of adaptive courses. We used the SUS in a two-factor structure with usability and learnability as factors and compared the results of adaptive courses to the results of non-adaptive courses.

Section II presents background information about the SUS from the literature and findings from our previous research on system usability scale with non-adaptive courses. Section III contains a statistical analysis of data and comparison with previous findings. Section III also includes analyses with two additional items for the learnability factor. Section IV discusses findings from both studies. Section V comprises the conclusion and future work possibilities to improve our findings.

II. BACKGROUND INFORMATION AND LITERATURE

Usability of web applications and especially learning management systems is a popular research area [3][4]. There are different scales for measuring usability such as the Computer System Usability Questionnaire (CSUQ) [5], the Questionnaire for User Interface Satisfaction, (QUIS) [6] or the Software Usability Measurement Inventory, (SUMI) [7], which contain more questions than the SUS and are more complex to analyze. Translations of the SUS can be found in different languages, including Spanish, French, Greek, and German. In our previous study, we applied and validated a German version of the SUS.

The original SUS is a one-dimensional tool aiming to measure usability. Although not used widely, there are studies that use a two-factor structure, mainly on websites [8][9]. This two-factor structure was tested and validated for the German version of the SUS in an e-learning domain with our previous study [2].

Digital learning systems are considered adaptive if they can dynamically change to better suit the learning in response to information collected during the course of learning rather than on the basis of preexisting information, such as learner's age, gender, or achievement test score, as defined by the U.S Department of Education, Office of Educational Technology [10]. However, digital learning systems can also integrate preexisting information such as gender or the results of entrance tests. It is possible to list three main groups as a basis for adaptation: (1) gender, learning style, or emotions can be grouped as personal

 ISBN: 978-1-61208-754-2

characteristics, (2) topics or task difficulty can be divided as content-specific characteristics, and (3) learning time or place can be grouped as context-based characteristics [11]. These groups enable adaptation in different dimensions throughout the learning experience [12].

Task difficulty, actual knowledge prior to taking the course and adaptive support (hints, feedbacks, level of detail) were used in our adaptive courses. The online learning platform used in our study was Moodle, as this is the learning management system used at our university. Adaptive courses in Moodle were different from classical courses as they contained more interactive tasks, specific feedbacks based on learner`s responses and different learning paths for each learner. Depending on the result in a task, a new task was recommended to the learner. The learner was not expected to complete all tasks to be prepared for the exam at the end of the course.

On the other hand, classical courses in Moodle contained the same tasks and learning content for all learners, in the same order, disregarding their previous knowledge level, performance, and progress during the course. Non-adaptive courses proposed the same path for all learners, and learners did not receive any feedback nor recommendations about what to study next.

In our previous study, data from 722 students, enrolled in non-adaptive courses in Moodle, were collected and analyzed in two different semesters (211 students in spring semester 2015 and 511 students in autumn semester 2015/16). Table I summarizes the findings from these two data sets.

TABLE I. FINDINGS FROM PREVIOUS STUDY

N	SUS Score[a]	SD	α	Item Discrimination power	Difficulty
211	62.87	21.74	0.91	0.54 - 0.79	0.58 - 0.67
511	67.51	19.55	0.90	0.51 - 0.79	0.60 - 0.73

Legend: N = number of participants, SD = standard deviation, α = Cronbach's Alpha

a. SUS Score of 10 Items

Exploratory factor analysis of the first data set (n=211) led to a two-factor structure, consistent with previous models [8][9]. The usability factor was formed with the items 1, 2, 3, 5, 6, 7, 8, 9 and the learnability factor with the items 4 and 10. In our previous study, after further confirmatory factor analysis, using Amos [21], of the second data set (n=511) items 5 and 6 were removed due to high residual values. This current study aims to compare results from adaptive courses with the results from the non-adaptive courses.

III. METHODOLOGY AND RESULTS

This section contains statistical analysis of data and compares results from adaptive and non-adaptive courses.

A. *Data*

The current study contains data from students enrolled in courses on Mathematics, Statistics, and Science. The evaluation of the adaptive courses with the SUS was carried out together with the standard module evaluation that takes place at the end of each course. Module evaluation questionnaire included questions about the course content and the performance of the lecturer. Students were asked to complete the SUS in order to measure the usability, as of the module evaluation questionnaire focused more on the module content rather than the usability of the online-course. To increase completion rate students were asked to take the module evaluation and therefore the SUS during the last classroom unit.

271 students received the survey and 118 of these students completed and returned it, resulting in a high completion rate of 43%. All courses were adaptive and contained pre-knowledge tests and tasks in detailed and non-detailed versions. After each task, the system recommended a new task to the students. All these courses were taught in a blended learning environment, with 80% self-study rate (including online activity using Moodle) and 20% classroom presence with face-to-face interaction with the lecturer and other students.

All online SUS-questionnaire items could be answered on a Likert scale ranging from 1="does not apply" to 5="is absolutely true". Half of the items had an inverted scale. To correctly calculate the SUS value as suggested by Brooke [1] the scale had to be converted from 1-5 to 0-4 and the directionality of the scales had to be consistent. For this, one was subtracted from the value for each odd numbered question, "value - 1", and the value for each even numbered question was subtracted from 5, "5 - value" [1]. The sum of the values for all 10 items was then multiplied by 2.5 resulting in the total SUS score ranging from 0 to 100. The average SUS score for was 55.08 with a standard deviation of 20.20.

B. *Analysis of Data*

This next part contains the reliability analysis of the scale. Analyses were done using R [13]. The "psych" package [14] was used to check reliability, validity and exploratory factor analysis and the "lavaan" package [15] was used for confirmatory factor analysis.

Cronbach's alpha value is a good measure to estimate internal consistency and therefore reliability of the scale. Cronbach's alpha value was 0.91 for 10 items, which is well above the widely used limit of 0.7 [16]. Cronbach's alpha values for the two factors, 0.91 for usability factor (items 1, 2, 3, 5, 6, 7, 8, 9) and 0.81 for learnability factor (item 4 and 10) were also above 0.7.

Item discriminatory power and item difficulty were used to assert item validity. In order to evaluate item discriminatory power of our scale, we applied the corrected item total correlation test. Corrected item total correlation values were calculated to be between 0.56 and 0.91. These values are acceptable, based on the common assumption of being greater than 0.3 [17]. To calculate item difficulties for each item, we divided the average response for each question with the maximum possible value (which was 5 for our scale). The difficulty of the questions lay between 0.55 and 0.69, which is in the acceptable range of 0.20 to 0.80 [18].

The Kaiser-Meyer-Olkin test was used as a measure of how suitable our data was for factor analysis. A Kaiser-Meyer-Olkin test value of 0.89 was calculated for our study, which is defined as “meritorious” by Kaiser and Rice [19], implying that our survey data was suitable for factor analysis.

Exploratory factor analysis performing maximum-likelihood method supported the two-factor structure with usability and learnability as factors. The usability factor had sums of squared loading of 3.745 and the learnability factor had sums of squared loadings of 2.485. The two factors explained the variance at 62.3%. The results of the exploratory factor analysis signaled a problem in some items (Item 2, Item 6, and Item 8). These items had similar loadings for both factors and could not clearly be classified as belonging to either of the factors. Although the factor structure was supported, the difference between factor loadings were very small for Item 2, Item 6 and Item 8 had 0.071, 0.051 and 0.052 respectively. After checking these items in detail, we found that the wording of the German translation might have been unclear (in the context of adaptive courses) and might have affected students’ responses. In the discussion, we will be reflecting on this.

We performed confirmatory factor analysis, keeping this in mind and calculated the values separately for the complete 10 items and the remaining 7 items after excluding items 2, 6, and 8. Removing the three items, increased the Confirmatory Fit Index (CFI) (from 0.896 to 0.966) and the Tucker-Lewis Index (TLI) (from 0.863 to 0.946) and decreased the Root Mean Square Error of Approximation (RMSEA) (from 0.136 to 0.097) and the Standardized Root Mean Square Residual (SRMR) (from 0.073 to 0.05). These values were in compliance with guidelines to have greater CFI and smaller RMSEA and SRMR measures [20]. Removing these items clearly improved the findings from confirmatory factor analysis. Confirmatory factor analysis was also done with a single factor structure to further test if two factor structure performs better, which resulted in a low CFI (0.837) and high RMSEA (0.168) and SRMR (0.09). These values confirmed that using a two-factor structure was suitable for this kind of usability test.

Table II shows the two-factor distribution, with and without items 2, 6 and 8, calculated with Varimax Rotation. When items 2, 6 and 8 were removed, total variance explained decreased to 47.1% composed of 3.043 for usability and 1.666 for learnability (sums of squared loadings). German translation of the SUS items and the original versions in English can also be found in Table II.

In order to check reliability, we recalculated Cronbach’s alpha values for all 7 items, 5 usability items and 2 learnability items. These values were found to be 0.86, 0.89 and 0.81 respectively, confirming reliability.

C. Comparison of Adaptive and Non-adaptive Courses

In this part, our current results will be compared to the results from the second data set (n=511) of our previous study. As the confirmatory factor analysis was implemented with the second data set only, we used this data set for the comparison.

The overall SUS score for the original 10 items were 55.08 for the adaptive courses (current study) and 67.51 for the non-adaptive courses (previous study).

TABLE II. EXPLORATORY FACTOR ANALYSIS

Item Nr.	Item	10-Item		7-Item	
		F #1	F #2	F #1	F #2
1	Ich würde gerne häufiger Module [a] wie dieses besuchen. I think that I would like to use this system frequently.	**0.78**	-0.09	**0.78**	-0.08
2	Dieses Modul war unnötig kompliziert. I found the system unnecessarily complex	-0.49	0.56		
3	Es war einfach mit diesem Modul zu lernen. I thought the system was easy to use.	**0.78**	-0.29	**0.83**	-0.22
4	Ich brauchte Support fürs Lernen in diesem Modul. I think that I would need the support of a technical person to be able to use this	-0.23	**0.77**	-0.31	**0.65**
5	Die verschiedenen Aktivitäten waren in diesem Modul gut integriert. I found the various functions in this system were well integrated.	**0.84**	-0.21	**0.84**	-0.14
6	Es gab zu viele Ungereimtheiten in diesem Modul. I thought there was too much inconsistency in this system.	-0.42	0.47		
7	Die meisten würden mit diesem Modul sehr schnell zurechtkommen. I would imagine that most people would learn to use this system very	**0.70**	-0.27	**0.70**	-0.26
8	Es war sehr mühsam mit diesem Modul zu lernen. I found the system very cumbersome to use.	**-0.67**	0.62		
9	Ich fühlte mich in diesem Modul sehr sicher. I felt very confident using the system.	**0.64**	-0.36	**0.68**	-0.34
10	Es brauchte viel Vorarbeit, bevor ich mit diesem Modul lernen konnte. I needed to learn a lot of things before I could get going with this system.	-0.08	**0.80**	-0.09	**0.99**

a. Courses are named as “Modules” in our university.

To be able to compare the SUS values with our altered item numbers we standardized the results to be in the same range as the original SUS, namely to be between 0 and 100. Table III below presents these "normalized SUS scores" for learnability (items 4 and 10) and usability (items 1, 3, 5, 7, 9) factors. It is worth noting that, in the previous study, item 5 was removed from the model together with item 6 due to high residuals, but in order to make comparison easier we included item 5 in the Table III results. We observe that, scores for new data set, which is for adaptive courses are lower than that of the previous data set for non-adaptive courses for both usability and learnability related items with similar standard deviations.

TABLE III. FACTOR BASED SUS SCORES

Items for Calculation	Old Data (N=511)		New Data (N=118)	
	SUS	SD	SUS	SD
2 Items (Learnability)	62.13	27.01	48.41	25.92
5 Items (Usability)	68.47	20.00	57.80	21.67

For the confirmatory factor analysis, we compared results from our current study (n=118) to our previous study (n=511). Our previous study reported slightly higher CFI (0.94 versus 0.966) and lower RMSEA (0.070 versus 0.097) values after excluding the low-loading items. Table IV summarizes these findings from confirmatory factor analysis.

TABLE IV. COMPARISON OF CONFIRMATORY FACTOR ANALYSIS

Data Set	Confirmatory Factor Analysis Results			
	N	*Remarks*	*CFI*	*RMSEA*
1	511	8 Items (5 and 6 removed)	0.94	0.070
2	118	9 Items (2, 6 and 8 removed)	0.966	0.097

D. SUS with 12-Items

The factor learnability has only two items in the original version of the SUS. In order to strengthen this factor and for a better explanation in the variance, we formulated two additional questions and added them as listed below. Table V presents the new items in English and German.

TABLE V. NEW ITEMS FOR LEARNABILITY FACTOR

Item	Language	New Item
Item 11	German	Ohne Unterstützung (von Kommilitonen, Dozierende etc.) hätte ich diesen Online-Kurs nicht verstanden.
	English	Without support (from fellow students, lecturers etc.), I would not have understood this online course.
Item 12	German	Ich hatte immer wieder Fragen bzgl. dieses Online-Kurses.
	English	I frequently had questions about this online course.

Although there are several definitions of learnability, the improvement in performance after repeated trials can be taken as a simple definition. A good learnability should lead to a high level of proficiency of the user, within a short time and with minimal effort [22]. The original SUS scale has only two items to measure learnability, item 4 states support from technical staff needed and item 10 addresses one`s own effort to be able to use the system.

Newly proposed item 11 "*Without support (from fellow students, lecturers etc.) I would not have understood this online course*", points to support from other users including fellow students and lecturers, but not technical experts, which implies extensive usage and increasing time spent with the system enables enough knowledge to support others.

Item 12 "*I frequently had questions about this online course*" focuses on learner`s reflection about the system and having questions regarding the system as a dimension of learnability.

As a first step of analysis, the reliability of the survey with the added questions was tested. Cronbach`s alpha value was calculated for all 12 items, the remaining 9 items after removing items 2, 6 and 8, the (remaining) 5 items for the usability factor and the 4 items for the learnability factor. Table VI shows these values in addition to item discriminatory power and item difficulty. We applied the same calculation as in part "C" to calculate the "normalized SUS score" for the learnability factor with 4 items (44.76±25.04), which is slightly lower than the value presented in Table III (48.41±25.92) for the "normalized SUS scores" of usability and learnability factors for 511 and 118 participants.

In a next step, exploratory and confirmatory factor analyses including the new items were conducted. All items showed clear distributions in factor loadings (Table VII) and appropriate CFI (0.964), RMSEA (0.084) and SRMR (0.05) values.

TABLE VI. COMPARISON OF QUALITY CRITERIA

Included Items	α	Item discriminatory power	Difficulty
12 Items	0.92	0.59 - 0.9	0.52 - 0.69
9 Items (2,6, 8 excluded)	0.89	0.58 - 0.79	0.52 - 0.69
5 Items (Usability)	0.89	0.74 - 0.83	0.65 - 0.69
4 Items (Learnability)	0.87	0.69 - 0.89	0.52 - 0.62

The usability factor had sums of squared loadings of 3.116 and the learnability factor 2.636, and both factors explained 57.5% of the variance. Table VII presents the factor loadings with the remaining 9 items after removing items 2, 6 and 8 and adding items 11 and 12. The usability factor contains items 1, 3, 5, 7 and 9 and the learnability factor contains items 4, 10, 11 and 12.

TABLE VII. FACTOR LOADINGS AFTER REMOVAL OF ITEMS

Item	Usability	Learnability
1	**0.782**	-0.107
3	**0.798**	-0.312
4	-0.207	**0.913**
5	**0.816**	-0.216
7	**0.724**	-0.19
9	**0.678**	-0.276
10	-0.138	**0.698**
11	-0.261	**0.731**
12	-0.299	**0.717**

Results presented in the tables above, show that the addition of two new questions for the learnability factor increased overall performance of the SUS tool by increasing

 ISBN: 978-1-61208-754-2

the total variance explained to 57.5% (31.1% from usability and 26.3% from learnability) from 47.1% (30.4% from usability and 16.6% from learnability). In addition, the RMSEA value was reduced from 0.097 to 0.084, causing almost no change in the CFI (0.966 to 0.964) and SRMR (0.05 for both cases) values.

IV. Discussion of the Results

This study uses a German version of the System Usability Scale, applied to an e-learning content with adaptive courses and compares the results to previous findings for non-adaptive courses. The reliability and validity of the German version of the SUS are intact for both the current and the previous study. This also holds true when two newly created items are added to the scale in order to counterbalance the two-factor structure.

The application of a short questionnaire like the SUS in order to measure the usability of a learning portal has the potential shortcoming of not differentiating clearly between the usability and the content. It is hard for the students to answer some of the questions considering only the usability without including their experiences and feelings about the learning content. In order to make content and usability more distinguishable, a detailed module evaluation was directed to the students prior to the SUS. This module-evaluation contained questions about lecturer, course content, course literature, and classroom presence.

Exploratory factor analysis supported the two-factor structure of the SUS. This is in accordance with our previous study and further literature [6][7]. The exploratory factor analysis further uncovered 3 items with unclear factor loading, items 2, 6, and 8. This lead to an unacceptably high RMSEA value. After exclusion of these three items, the RMSEA was found to be in an acceptable range. The wording of these three items might have been unclear and might have led to some confusion in the assessment of the adaptive courses. For example Item 8: I found the system very cumbersome to use (in German “Es war sehr mühsam mit diesem Modul zu lernen”). We wanted to measure the usability of the module, but in this question, we did not clearly separate the learning content (e.g. the adaptive online tasks) of the module from the usability of the module. Despite answering the detailed module evaluation prior to the SUS questionnaire, students still might have assessed the complicated usability, the more complex online tasks, or even the complexity of the content. This could be a reason for which the factor load for this item is unclear. This supports the unsuitability of this item in our context. The same problem can also be seen in items 2 and 6.

To balance out the two-factor structure of the SUS we added 2 items meant to load onto the learnability factor. The new 9-item SUS (excluding items 2, 6, 8 and including items 11 and 12) had a clear distribution of factor loadings and CFI, RMSEA and SRMR very similar to the 7 item SUS (excluding items 2, 6, and 8) but had a higher explained variance (57.5% instead of 47.1%). These results validate the inclusion of the new items.

It is to be noted, that the total SUS value for adaptive courses was lower than the total SUS value for non-adaptive courses. This might be due to the richness or the voluntariness of the course. The adaptive courses include a variety of learning content, such as the previous knowledge tests, adaptive online tasks and various recommendations. Breaking down the total SUS value both factors were lower for the adaptive courses than for the one on non-adaptive ones. The difference in learnability factor is much bigger, which might mean, that students need more time to find their way around the course. In addition, students are free to use the adaptive learning system and following the recommendations. They are free to leave their adaptive learning path at any time. They do not have to follow the recommendations. All tasks can be accessed without limitations. Consideration of all these options can lead to an additional cognitive load for some students and therefore lead to a decrease in the usability of the course. We are conscious of this problem in usability and are constantly searching for possibilities to improve the course structure and try to present a simpler overview for adaptive courses. As part of this work (for example after changes in the course structure), we need a short questionnaire that students can answer to evaluate if and how the changes affected them. For this, we need the SUS with 12 items.

As the adaptive courses are quite complex a two factor (Usability and Learnability) structure for the SUS made sense. As with the total SUS value both factors were lower in the analysis of the adaptive courses than the one on non-adaptive ones. The difference in learnability factor is much bigger, which might mean, that students need more time to find their way around the course.

From the course-development side, we understand that the first orientation for students in adaptive courses requires more time. Students are accustomed to non-adaptive courses at the university and therefore, they have to adapt to the new adaptive course structure and hence sometimes have questions about how to work with the adaptive courses. We are working on making this transition seamless and rapid and on making adaptive courses more self-explanatory.

User experience was not measured in this current study (2018/19), nor was it measured in the reference study (2015/16). Therefore the user experience could not be compared. Future studies can take user experience into consideration, specially looking at improvements in user experience across studies.

German translation of some items (items 2, 6, and 8) were not clear and in the context of adaptive learning, could cause misunderstanding and misinterpretation. This problem in translation is an important limitation of this study, which should be considered in future studies.

V. Conclusion and Future Works

The blended learning offer of the examined university is based on a concept with around 20% interaction with a teacher (i.e in classrooms) and 80% self-learning time, which is guided and supported by the online learning system Moodle. The introduction of adaptive online courses was primarily designed to support the self-study phase of the students. In this context, the efficient use of the functionalities of the learning platform plays an important role with regard to a high degree of usability and learnability. This is especially important in order to prevent the students from a high cognitive load due to a difficult handling of the learning platform or within the courses, which would affect negatively the learning performance of the students. In fact, the SUS value (55.08) was considerably lower than that of the non-adaptive courses examined (62.87 and 67.51, respectively). This is a distinct indication that the usability of adaptive courses should be increased. If we consider the SUS value of usability (57.8) and learnability (48.41 respectively 44.76) we see, that the value for learnability is much smaller. It takes a longer time for a student to understand the orientation and flow of the course in their first appearance in an adaptive course. We are currently working on improving the clarity of the adaptive online courses and we have started to produce short explanatory videos, explaining how to learn with adaptive courses. We aim to make the process of introduction (learnability) to adaptive course structure easier with these explanatory videos.

The limits of the short scale were also shown here. Accordingly, after calculating the SUS values, we propose to examine more closely how the students have used the course and what progress they have made. We suggest that we have already done this in previous studies, to analyze the log files from the Learning Management Systems database more closely and to relate them to indicators of the students' learning progress. This form of analysis has the advantage that objective data was collected without interfering with the students' learning process, for example by asking questions. This type of data collection also made it possible to efficiently evaluate large datasets or datasets over a longer period of time. In individual cases, these forms of data collection could be supplemented by questions or interviews. To sum up, the SUS has proven to be a reliable instrument also for adaptive courses. However, it also turned out that its German version still has some weaknesses, as demonstrated by the linguistic problems and the exclusion of items. The next step is to test the original scale again in larger samples and test it for potential effecting variables, such as experience with Moodle or satisfaction with IT support or classroom sessions. In our previous study [2], we found that satisfaction with classroom sessions or the lecturer has an influence on the SUS value. These factors may also be important in adaptive online courses.

REFERENCES

[1] J. Brooke, SUS: a “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & A. L. McClelland (eds.), Usability evaluation in industry (pp. 189-194). London: Taylor and Francis, 1996.

[2] P. Bergamin, E. Werlen, M. Holthaus, and M. Garbely, System Usability für E-Learning-Anwendungen "Validierung einer deutschsprachigen Version einer Kurzskala zur Messung der Usability in einem Blended Learning Szenario", unpublished.

[3] R. P. Bringula, “Influence of faculty-and web portal design-related factors on web portal usability: A hierarchical regression analysis”, Computers & Education, 68, pp. 187-198, 2013.

[4] N.J. Navimipour and B. Zareie, “A model for assessing the impact of e-learning systems on employees’ satisfaction”, Computers in Human Behavior, 53, pp. 475-485, 2015 .

[5] J. R. Lewis, "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use", International Journal of Human-Computer Interaction, 7(1), pp. 57-78, 1995.

[6] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the humancomputer interface", In Proceedings of the SIGCHI conference on Human factors in computing systems ACM Press, New York, pp. 213–218, 1988.

[7] J. Kirakowski and M. Corbett, "SUMI: The software usability measurement inventory", British journal of educational technology 24, 3, pp. 210–212, 1993.

[8] S. Borsci, S. Federici, and M. Lauriola, "On the dimensionality of the System Usability Scale: a test of alternative measurement models", Cognitive processing, 10(3), pp. 193-197, 2009.

[9] J. R. Lewis and J. Sauro, "The factor structure of the System Usability Scale", In Proceedings of the 1st International Conference on Human Centered Design: Held as Part of HCI International 2009, pp. 94-103. Berlin, Heidelberg: Springer-Verlag, 2009.

[10] U.S. Department of Education, Office of Educational Technology, 2013.

[11] K. Wauters, P. Desmet, and W. van Den Noortgate, Adaptive item-based learning environments based on the item response theory: Possibilities and challenges. In Journal of Computer Assisted Learning, 26(6), pp. 549-562, 2010.

[12] M. Holthaus, F. Hirt, and P. Bergamin, "Simple and Effective: An Adaptive Instructional Design for Mathematics Implemented in a Standard Learning Management System," CHIRA 2018 - 2nd International Conference on ComputerHuman Interaction Research and Applications - Proceedings, pp. 116 - 126, 2018.

[13] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ [Accessed: 01-Oct-2019]

[14] Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych [Accessed: 01-Oct-2019]

[15] Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36, http://www.jstatsoft.org/v48/i02/ [Accessed: 01-Oct-2019]

[16] J. C. Nunnally, Psychometric Theory, 2nd ed., McGraw-Hill, New York, NY, 1978.

[17] E. Cristobal, C. Flavian, and M. Guinaliu, "Perceived e-service quality (PeSQ):measurement validation and effects on consumer satisfaction and web siteloyalty", Managing Service Quality 17 (3), pp. 317–340. 2007.

[18] H. D. Mummendey and I. Grau, "The questionnaire method: basics and application in personality, attitude and selfconcept research", Hogrefe Publishers, 2014.

[19] H. F. Kaiser and J. Rice, "Little Jiffy, Mark Iv". Educational and Psychological Measurement, 34, pp. 111-117, 1974.

[20] D. Hooper, J, Coughlan, and M. Mullen, "Structural Equation Modelling: Guidelines for Determining Model Fit", Electronic Journal of Business Research Methods, 6(1), pp. 53-60, 2008.

[21] J. L. Arbuckle, Amos (Version 23.0) [Computer Program]. Chicago: IBM SPSS, 2014.

[22] M. Unsöld, "Measuring Learnability in Human-Computer Interaction", Masters thesis, Ulm University, 2018.