ACTA INFOLOGICA 2022;6(1):19-32
dergipark.org.tr/acin

DOI: 10.26650/acin.1077400

RESEARCH ARTICLE

Usability Evaluation of the Online Skill Assessment Tool

Online Beceri Değerlendirme Aracının Kullanılabilirliğinin Değerlendirilmesi

Merve Yıldız[1], Muhammet Berigel[2], Furkan Kalyoncu[3], Özlem Özgenç[4]

[1] (Lect. PhD.), Karadeniz Technical University, Department of Distance Education Application and Research Center, Trabzon, Turkiye
[2] (Assoc. Prof.), Karadeniz Technical University, Department of Management Information Systems, Trabzon, Turkiye
[3] (Expert), Karadeniz Technical University, Department of Technology Transfer Application Research Center, Trabzon, Turkiye
[4] (Assoc. Prof.), Karadeniz Technical University, Department of Forest Products Engineering, Trabzon, Turkiye

ORCID: M.Y. 0000-0002-8665-0287;
M.B. 0000-0001-5682-8956;
F.K. 0000-0003-2214-3347;
Ö.Ö. 0000-0001-9428-5681

Corresponding author:
Merve YILDIZ
Karadeniz Technical University, Department of Distance Education Application and Research Center, Trabzon, Turkiye
E-mail address: mrvyildiz.86@gmail.com

Submitted: 22.02.2022
Revision Requested: 18.05.2022
Last Revision Received: 27.05.2022
Accepted: 27.05.2022
Published Online: 09.06.2022

Citation: Yildiz, M., Berigel, M., Kalyoncu F., and Ozgenc, Ö. (2022). Usability evaluation of the online skill assessment tool. *Acta Infologica, 6*(1), 19-32. https://doi.org/10.26650/acin.1077400

ABSTRACT

This study was conducted to evaluate the usability of the online skill assessment tool which was developed to evaluate the literacy, numeracy, and digital skills of adults with low skill levels. The sample consisted of 14 participants who were between the ages of 18 and 40 and who had graduated from secondary school at most. In the study, the user-based method, in which user experiences are employed, was used, and according to the effectiveness, efficiency and satisfaction criteria determined by ISO 9241-11, the usability of the online skill assessment tool was evaluated. According to the data obtained from the screen recordings, effectiveness was considered as the completion rate of the tasks, and efficiency was considered as the completion duration of the tasks. Satisfaction was also considered as the values obtained from the questionnaire data. The findings indicated that the participants were able to complete most of the tasks and were generally satisfied with the use of the system. Some usability problems were identified as a result of the observations and the detailed examination of the screen recordings. These problems have been a guide for updates to render the system more usable before the system is used by end users.

Keywords: Usability, Online Assessment System, Skill Assessment, User-based Method

ÖZ

Bu çalışmada, düşük beceri seviyesine sahip yetişkinlerin okuryazarlık, sayısal ve dijital becerilerinin değerlendirilebilmesi için geliştirilen online beceri değerlendirme aracının kullanılabilirliğinin değerlendirilmesi amaçlanmıştır. Bu amaç doğrultusunda, çalışma grubunu, 18-40 yaşları arasında ve en fazla ortaokul mezunu olan 14 katılımcı oluşturmaktadır. Çalışmada, kullanıcı deneyimlerinin işe koşulduğu kullanıcı tabanlı yöntem kullanılmış ve online beceri değerlendirme aracının kullanılabilirliğinin değerlendirilmesinde, ISO 9241-11 tarafından belirlenen etkililik, verimlilik ve memnuniyet ölçütleri temel alınmıştır. Bu bağlamda, görevlerin tamamlanma oranı olarak kabul edilen etkililik ile görevlerin tamamlanma süresi olarak kabul edilen verimlilik ekran kayıtlarından elde edilen verilere göre belirlenmiştir. Memnuniyet ise araştırmacılar tarafından geliştirilen anketten elde edilen veriler doğrultusunda belirlenmiştir. Elde edilen bulgular, katılımcıların, görevlerin çoğunluğunu tamamlayabildiklerini ve sistemin kullanımı ile ilgili genel olarak memnun olduklarını göstermiştir. Bunun yanı sıra, kullanılabilirlik testi sırasında yapılan gözlemler ve ekran kayıtlarının detaylı incelenmesi sonucunda bazı kullanılabilirlik problemleri tespit edilmiştir. Bu problemler, sistemin son kullanıcılara açılmadan önce, daha kullanılabilir olması adına sistemin tasarımında yapılacak güncellemeler için yol gösterici olmuştur.

Anahtar kelimeler: Kullanılabilirlik, Online Değerlendirme Sistemi, Beceri Değerlendirme, Kullanıcı Tabanlı Yöntem

1. INTRODUCTION

Based on the development of information and communication technologies, access to digital systems such as web pages, online environments and mobile applications has become easier. Digital systems are preferred more by people as they provide a more comfortable and flexible usage opportunity (Anam, Sadiq, & Jamil, 2020). However, users who easily access a system and start using it quickly may leave the system with the same speed when they cannot use the system (Ersoy, 2004; KAMIS, 2018). This is because people have differences in terms of gender, age, nationality, culture, experience, skill, or disability. All these individual characteristics determine whether a person will use that system or not (Çağıltay, 2016; Dix et al., 2004; Şahin, Hebebci, & Çelik, 2014). Therefore, it is an important issue to design systems acceptably so that everyone can use them (de Castro Lozano et al., 2011; Newell & Gregor, 2002). At this point, the concept of universal design comes in sight. Universal design is defined as a product design process that can be used by as many people as possible in as many situations as possible (Dix et al., 2004). Story, Mueller, & Mace (1998) proposed seven universal design principles for products or systems to be designed: equitable use, flexibility in use, simple and intuitive, perceptible information, tolerance for error, low physical effort, size and space for approach and use. These principles provide a framework to develop a user-centered design for researchers and developers. Certainly, this does not mean that there will be no problems in a well-designed system. Thus, systems that are developed need to be tested. At this point, the concept of usability also comes to the fore.

Usability can be defined as a combination of factors that affect the user's interaction with the product or system (Nielsen, 1994). According to the definition of ISO 9241-11 (1998), usability is "the extent to which a product (hardware, software or service) can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". In this context, usability should be evaluated to determine the usability measure of a system or product. There are many different methods and approaches for evaluating usability. Dillon (2001) grouped evaluations of usability under three categories: user-based, expert-based, and model-based. In the user-based method, an application is tested with a sample of users performing a set of pre-determined tasks. The expert-based method refers to any form of usability evaluation which involves a group of HCI experts examining the application and estimating its likely usability for a given user population. In model-based approaches, an HCI expert employs several methods to predict certain aspects of user performance. This approach involves measuring only one component of usability and limited task applicability. In the literature, it is stated that the user-based method is superior to the others in terms of generally providing reliable and valid estimates in usability evaluation (Dillon, 2001), finding deep usability problems (Lyzara et al., 2019) and being effective in exploring various interface options at the early stages of system design (Ogolla, 2011).

Usability is one of the most important features that determine the quality of a product or system. Unusable products and systems lead to lower efficiency and effectiveness, and so, end users face difficulties while using them (Sagar & Saha, 2017). This also causes the product or system to fail in real-world practice (Seffah et al., 2006). Thereby, there are many usability studies in various fields for different purposes. When the literature is examined, it is seen that usability studies have generally focused on websites. There are examples of studies on the website usability of a university (Akgül, 2020; Cengiz, 2016; Hasan, 2014; Ismail & Kuppusamy, 218; Mentes & Turan, 2012; Pant 2015; Van Den Haak, De Jong, & Jan Schellens, 2003), e-commerce (Gonçalves et al., 2018; Hasan, Morris, & Probets, 2012; Lee & Koubek, 2010; Li & Li, 2011), e-government (Chang & Almaghalsah, 2020; Huang & Benyoucef, 2014; Paul & Das, 2019) and many more. Alongside these, evaluation of the usability of educational environments such as e-learning platforms (Granić & Ćukušić, 2011; Harrati et al., 2016; Shi et al., 2013; Ventayen et al., 2018), learning management systems (Althobaiti & Mayhew, 2015; Kakasevski et al., 2008; Phongphaew & Jiamsanguanwong, 2017) and other tutoring systems (Chao et al., 2016; Lin et al., 2011; Pensabe-Rodriguez et al., 2020) is also within this field of study.

One of the most important dynamics of educational processes is assessment. According to the purpose of the assessment, it is divided into three categories - diagnostic, formative and summative (Dumit, 2012; Güven, 1987; Senemoğlu, 2013). In diagnostic assessment, it is aimed to determine the readiness levels, interests, tendencies, knowledge, and skill levels of students and direct them to the areas where they can be productive. The main purpose of formative assessment is to complete learning deficiencies by identifying learning deficiencies in the learning-teaching process and eliminating learning difficulties.

Summative assessment is usually made to determine the level of the student at the end of the learning-teaching process. In other words, it is aimed to reveal the progress made in line with the objectives of a course. Within this context, in the literature, it is possible to encounter studies on evaluation of the usability of various assessment systems developed for different purposes (Bayrak & Akcam, 2015; Battal & Çağıltay, 2015; Karahoca, Karaoca, & Günoğlu, 2009; Lai, Chen, & Chou, 2017; Majzub, 2009; Vairamuthu & Anouncia, 2016), even though the number of such studies is limited. However, when these studies were examined, it was seen that survey data mostly based on self-reported user opinions were used in the usability evaluation of assessment systems. Therefore, it may be stated that there is a gap in the literature regarding user-based usability evaluation of assessment systems. From this point of view, this study aimed to evaluate the online skills assessment tool which was developed to evaluate the literacy, numeracy, and digital skills of low-skilled/underqualified adults within the scope of the online skill assessment project, by using the user-based method.

2. ONLINE SKILL ASSESSMENT PROJECT

According to 2019 data from the Turkish Statistical Institute (TÜİK, 2019), 15% of secondary school graduates over the age of 15 were at that time unemployed in Turkey. For low-skilled/underqualified adults to contribute to the economy by entering the labor market, there is a need to provide opportunities to increase their skill levels with appropriate training. However, in order to be able to create appropriate training opportunities, it is necessary to determine the skills and the skill levels of these adults by first making diagnostic assessment. Within this context, in the project supported by the European Union Programme for Employment and Social Innovation "EaSI", it was aimed to determine the literacy, numeracy and digital skill levels of low-skilled/underqualified adults in Turkey. The objectives of the project are as follows:

- The main objective of the project is to develop a new, web-based and adaptive assessment tool to assess the literacy, numeracy and digital skills of low-skilled/underqualified adults and to test the reliability and validity of this tool.
- The second objective of the project is to present the evaluation results through the reports generated by the assessment tool to be developed. These reports will include information about both the current skill levels of individuals and the potential skills that individuals need to be improved.
- The final objective of the project is to provide the basis for creating tailored learning methods for adults.

In line with the main purpose of the project, an adaptive, web-based assessment tool which was named the "Online Skill Assessment Tool (OSAT)" was developed for assessment of skill levels. The target group with low-skilled/underqualified individuals consisted of adults between the ages of 18 and 40 who had graduated from secondary school at most. Skills were determined as Turkish literacy, numeracy and digital skills. Question pools were created for the three skill areas that were determined. A total of 1053 questions, 338 of which are about Turkish literacy, 303 of which are about numeracy and 412 of which are about digital skills, were prepared by field experts. The questions were prepared in accordance with a conceptual framework based on the European Qualifications Framework (EQF). The system works according to the CAT (Computerized Adaptive Testing) algorithm. CAT is a form of computer-based test that adapts to the examinee's ability level. In other words, it is a form of computer-administered test in which the next item or set of items selected to be administered depends on the correctness of the test taker's responses to the most recent items administered. In this direction, the developed system determines a different starting level for each user according to some demographic information that is entered. According to the adaptation in the system, each user is faced with questions of different difficulty levels. Depending on the answers to the questions, the questions get more difficult, or the test is interrupted.

Developed as an adaptive web-based assessment system, OSAT consists of two main modules which are the admin module and the application module. The admin module is the interface where authorized institutions that will use the system perform operations such as user management, test creation, question adding and reporting. The application module is the web interface where OSAT is located and can be used by the target group. If the user logs in to the application module for the first time, they must first register to the system. After the registration, the demographic form field appears in front of the user who logs into the system. After filling in the demographic form, the user has to choose institution information. After that, the user can access the test or tests defined for themselves by the institution from the test list in the menu. The user logs out of the

system after the relevant test or tests are completed. The tests consist of question types such as multiple choice, fill-in-the-blank, matching, true and false, but there are also questions with different features such as simulation, picture or audio questions in the question pool. In this context, while determining usability tasks for the application module for which a usability test was performed, in order to make a comprehensive evaluation, all menu operations and test interfaces of the application module were aimed to be included.

3. METHOD

3.1. Usability Evaluation Process of OSAT

In this study, the user-based method, in which user experiences are employed, was used to evaluate the usability of OSAT. In the process of the usability evaluation of OSAT shown in Figure 1, the usability framework determined by ISO 9241-11 (1998) was taken as a basis.

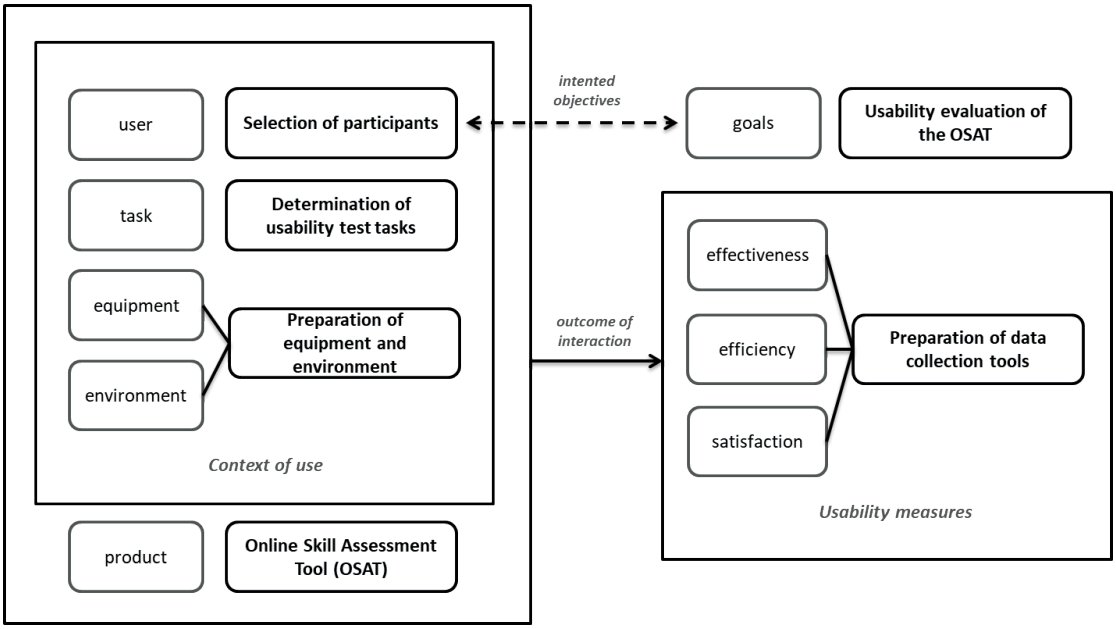

Figure 1. Usability evaluation process of OSAT

3.1.1. Selection of participants

Since the target group of the project was adults between the ages of 18 and 40 and who had graduated from secondary school at most, the criterion sampling method, one of the purposive sampling methods, was used in determining the participants (Yıldırım & Şimşek, 2013). In order to reach participants with these criteria, an interview was held with the Trabzon Employment Agency, one of the stakeholders of the project. The participants who were contacted were given information about the purpose and scope of the study. Accordingly, 18 participants agreed to participate in the study. However, four participants requested support from the researchers during the usability test, as their computer use abilities were low. Since there should be no intervention while performing the usability tests, the data obtained from these four participants were excluded from the analysis. Therefore, the study group consisted of 14 participants. Information on the demographic characteristics of the study group is presented in Table 1.

Table 1

Demographic information of the participants

		n
Gender	Male	8
	Female	6
Age	18-25	5
	26-30	5
	31-40	4
Education level	Graduated from secondary school	10
	Dropped out of high school	4
Employment status	Unemployed	6
	Permanent worker	4
	Temporary worker	3
	Self-employed	1

According to Table 1, eight men and six women participated in the study. Five of the participants were between the ages of 18 and 25, five were between 26 and 30, and four were between 31 and 40 years old. While ten of the participants had graduated from secondary school, four of them had dropped out of high school. Six participants were unemployed.

3.1.2. Determination of usability test tasks

Seven main tasks were determined for the usability test. In the sixth task, 14 sub-tasks related to each question type were determined. The descriptions of the tasks determined for the usability test are presented in Table 2.

Table 2

Usability test tasks

Task no	Task name	Task definition
1	Registration	Press the register link Fill in the registration form Press the send button
2	Login	Open the website Fill in the login form Press the login button
3	Demographic form	Click the menu Click the demographic form item Fill in the demographic form Press the save button
4	Institution information	Click the menu Click the institution information item Select the institution from the list Press the save button
5	Start test	Click the menu Click the test list item Select the test and click
6	Questions*	Answer the question Press the confirm and skip to next question button or Click the “I want to leave the question blank” box
6.1		Multiple choice (options text)
6.2		Multiple choice (options picture)
6.3		Multiple choice (audio question)
6.4		Short answer
6.5		More than one correct choice
6.6		Two options (true / false)
6.7		Two options (other)
6.8		Drag and drop into text
6.9		Short answer into text
6.10		Matching
6.11		Sorting
6.12		Simulation

6.13		Simulation
6.14		Simulation
7	Logout	Click the profile icon Press the logout button

** The same steps were followed for all question types (6.1-6.14)*

3.1.3. Preparation of data collection tools

Usability can basically be defined as effectiveness, efficiency, and satisfaction (ISO 9241-11, 1998). These three attributes common to evaluating usability are a measure of the impact on usability. At this point, these standard attributes may be defined as follows:

- Effectiveness: measures the degree to which users completely achieve specified goals correctly.
- Efficiency: measures the resources expended by users to completely achieve specified goals correctly.
- Satisfaction: measures which user needs are satisfied by using a product or system.

Depending on this, in the usability evaluation of OSAT, effectiveness; rate of task completion, efficiency; tasks completion duration, and satisfaction were considered as values obtained from the questionnaire data. Accordingly, the data collection tools are:

Screen recordings: Screen recordings were taken during the usability test in order to examine the behaviors of the participants on the system. And then, the task form was created to record information about the users' behaviors in a dataset in line with the tasks determined for the usability test. The data obtained from the screen recordings of each user, the task completion rate and other results for effectiveness and the task completion duration results for efficiency were entered in the task form.

Observation form: The observation form was developed to note the questions asked by the users during the usability test and important situations that could not be obtained from the screen recordings. The data obtained from the observation form were used to support the data obtained from the screen recordings.

Satisfaction questionnaire: The satisfaction questionnaire was developed to determine the opinions of the users about OSAT. Following a literature review, questionnaire questions on a 5-point Likert-type scale (1 = Strongly disagree to 5 = Strongly agree) consisting of 16 items were prepared (Appendix). Six of the items were about design, six were about navigation, and four were about ease of use.

3.1.4. Preparation of equipment and environment

The usability test was carried out at a meeting hall of the Eastern Black Sea Development Agency, which is the executive institution of the project. Before the implementation, the laptops on which the users would take the usability test were checked in terms of software and hardware. In order to prevent possible interruption problems, two different screen recording programs were installed on the laptops. Due to the Covid-19 pandemic, the headsets for audio questions and pens for filling in the forms were provided, one for each user. The data collection tools (observation form and satisfaction questionnaire), informed consent form and task list for the participants to be used in the data collection process were printed out. Finally, due to the Covid-19 pandemic, the necessary social distancing arrangement was created at the meeting hall, and disinfection procedures were carried out.

3.2. Data Collection Process

The participants contacted through the Trabzon Employment Agency, which is one of the associate organizations of the project, were notified via a message of the day and time of the usability test implementation. The data collection process was carried out in two sessions. Ten participants attended the first session, and eight participants attended the second session. Before the implementation started, after giving brief information about the project and the research process, the informed consent form was signed by the participants. After the screen recording programs were activated, OSAT's web page was

opened on the laptops. Important situations observed during the implementation were noted on the observation form by four researchers. The satisfaction questionnaire was given to the participants who then completed them. Later, incentive payments were given to every participant, and the implementation process shown in Figure 2 was completed.

3.3. Data Analysis

The data collected through screen recordings were analyzed, and the rate of completion and the completion duration for each task were reported. Mean and standard deviation were calculated for the data obtained from the satisfaction questionnaire. Additionally, usability problems related to OSAT were identified in line with the data obtained from the observation form and screen recordings. Frequency values were calculated for these usability problems.

Figure 2. Data collection process

4. FINDINGS

4.1. Findings on Effectiveness

In the usability evaluation of OSAT, effectiveness was handled as the rate of completion of the tasks. Correspondingly, the rate of completion of each task and the rate of each participant's task completion were calculated and are presented in Table 3.

Table 3

Findings on effectiveness

Task no	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Rate of each task completion
1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
4	√	X	X	X	√	X	X	X	X	√	√	√	X	√	6/14
5	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
6.1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
6.2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
6.3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
6.4	X	√	√	√	√	√	√	√	X	√	√	√	√	√	12/14
6.5	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
6.6	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
6.7	√	√	√	√	√	√	√	√	√	√	√	√	√	√	14/14
6.8	X	√	√	√	√	√	√	√	√	√	√	X	√	√	12/14
6.9	√	√	√	√	√	√	√	√	√	√	√	X	√	√	13/14
6.10	X	√	√	√	√	√	√	√	X	√	√	X	√	√	11/14
6.11	X	X	√	X	√	√	X	√	X	√	√	X	√	√	8/14
6.12	X	√	√	X	√	√	X	√	X	X	√	X	√	X	7/14
6.13	X	√	√	√	√	√	√	X	√	X	√	X	√	X	9/14
6.14	√	√	√	√	√	√	√	√	√	X	√	√	√	X	12/14
7	√	√	√	√	√	√	X	√	√	X	√	X	√	X	10/14

Rate of each participant' task completion	14/ 20	18/ 20	19/ 20	17/ 20	20/ 20	19/ 20	16/ 20	18/ 20	15/ 20	16/ 20	20/ 20	13/ 20	19/ 20	16/ 20

P: participant

As seen in Table 3, when the rate of completion of each task was examined, it was seen that most of the tasks could be completed at a high rate. It was determined that the most problematic tasks were 4, 6.11, 6.12 and 6.13. When the observation data and the detailed results of the screen recordings were examined, it was determined that the institution information task number 4 could not be completed due to the fact that the 'select the institution' item could not be found in the menu. The problem in task 6.11 regarding the sorting question was that the drag and drop pointer was not detected. In the simulation questions 6.12 and 6.13, the completion rate of these tasks was low, as it was not understood how to perform these tasks. The findings on usability problems are explained in detail in section 4.4.

4.2. Findings on Efficiency

In the usability evaluation of OSAT, efficiency was handled as the task completion duration. Correspondingly, the mean of each task's completion duration and the total duration were calculated and are presented in Table 4.

Table 4

Findings on efficiency

Task no	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**Mean of each task completion duration (≈)**
1	60	73	18	53	19	89	70	46	327	180	20	21	36	28	74"
2	3	49	37	10	46	12	60	104	26	197	20	83	30	75	54"
3	275	256	180	297	188	308	351	273	345	369	296	218	144	156	261"
4	62	0	0	0	145	0	0	0	0	17	61	200	48	63	43"
5	120	14	21	67	10	97	43	20	47	94	51	142	36	66	59"
6.1	33	16	16	14	10	42	19	16	17	23	12	13	14	24	19"
6.2	87	60	52	27	23	67	81	27	49	54	45	57	47	54	52"
6.3	65	42	18	26	15	19	31	29	43	42	86	32	40	14	36"
6.4	46	48	12	67	43	53	57	55	45	22	21	12	61	21	40"
6.5	28	29	33	34	13	12	38	30	42	23	40	17	28	19	28"
6.6	26	18	16	10	14	20	32	52	21	24	10	11	18	14	20"
6.7	46	40	35	40	23	49	30	25	45	22	20	32	28	35	34"
6.8	55	71	39	47	54	63	47	40	61	43	47	46	28	35	48"
6.9	28	38	20	46	23	44	45	34	40	50	33	29	37	32	36"
6.10	35	49	43	24	20	54	32	40	142	34	25	51	17	30	43"
6.11	8	33	25	27	42	24	4	29	43	38	19	42	16	14	26"
6.12	55	131	55	50	60	78	67	43	60	142	42	70	46	20	66"
6.13	25	33	25	14	28	25	78	51	67	0	54	28	18	0	32"
6.14	26	33	21	27	27	60	31	47	41	0	21	41	31	0	29"
7	9	12	27	15	16	63	23	249	42	0	21	332	17	0	59"
Total duration (≈)	18'	17'	12'	15'	14'	20'	19'	20'	25'	23'	16'	25'	12'	12'	

P: participant, ": second, ': minute

When the total durations shown in Table 4 were examined, it was seen that the participants completed all the tasks approximately between 12 and 25 minutes. The mean of the completion durations of each task varied. The task with the longest completion duration was the demographic form filling task, task no 3. The reason for this was that there were several form fields, and the participants tried to complete the task without filling all of these fields. Additionally, it was determined that the mean completion duration of the login task no 2, which was the relatively shortest task compared to the others, was also long. When the observation data and the detailed results of the screen recordings were examined, it was observed that the participants tried to enter the system without registration. These findings on the usability problems are explained in detail in section 4.4.

4.3. Findings on Satisfaction

In the usability evaluation of OSAT, satisfaction was handled as the values obtained from the questionnaire data. Correspondingly, the mean and standard deviation for each item were calculated, and they are presented in Table 5.

Table 5

Findings on satisfaction

Item no	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	$\bar{X}$	SD
1	5	4	5	5	4	4	4	2	5	4	2	3	4	1	3.71	1.27
2	5	4	5	5	4	4	4	4	5	4	5	5	4	1	4.21	1.05
3	5	4	5	5	3	4	4	4	5	4	4	5	5	5	4.43	0.65
4	5	4	5	5	1	4	4	4	5	4	5	5	5	4	4.29	1.07
5	5	4	5	5	1	4	4	4	4	4	3	5	5	4	4.07	1.07
6	5	4	5	5	1	4	4	4	4	3	4	3	2	4	3.71	1.14
7	5	4	5	5	3	4	4	4	4	3	5	5	5	3	4.21	0.80
8	5	4	5	5	5	4	4	4	4	4	4	5	5	4	4.43	0.51
9	5	4	5	5	5	4	4	4	4	3	5	5	5	5	4.50	0.65
10	5	4	5	5	5	4	4	4	4	3	5	5	4	4	4.36	0.63
11	5	4	5	5	5	4	4	4	4	4	4	5	5	3	4.36	0.63
12	5	4	5	5	5	4	4	4	5	4	5	3	3	5	4.36	0.74
13	5	5	5	5	5	3	4	4	5	4	5	5	5	4	4.57	0.65
14	2	4	5	5	5	3	4	4	2	3	5	5	5	3	3.93	1.14
15	4	4	5	5	4	3	4	4	5	3	4	5	5	5	4.29	0.73
16	4	3	3	4	3	4	4	4	4	2	5	5	4	5	3.86	0.86

P: participant

As seen in Table 5, when the mean values were analyzed, it may be stated that the participants were generally satisfied with OSAT. The item no 13 which was expressed “generally, the system is easy to use”, supported this with the highest mean value. The items with the lowest mean values were item no 1-”the font and font size in the system make it easy to read”, and item no 6-”the colors of the navigation elements (such as buttons, scroll bars, drag-and-drop markers) are appropriate”. During the observations, it was determined that the question root statement was not seen due to the small font size, especially in the simulation questions, and therefore, the questions were skipped before they could be answered. Furthermore, the low mean value in item 6 coincided with the “drag and drop pointer was not detected” finding in section 4.1.

4.4. Findings on Usability Problems

In the usability evaluation of OSAT, some usability problems were identified as a result of the observations and screen recordings. These problems are presented in Table 6.

Table 6

Findings on usability problems

Task no	Problem	f
1 - 2	Trying to login to the system without registration	11
3	Trying to complete the task without filling all the demographic form fields	12
4	Cannot find institution information selection menu	10
5	Cannot find the test list in the menu	8
6.11	Cannot detect the drag and drop pointer in sorting question	5
6.12 - 6.13 - 6.14	Cannot understand how to respond to simulation questions	9
7	Cannot find the logout menu	6

As seen in Table 6, one of the situations where the participants experienced the most problems was trying to login to the system without registration. It was determined that this was due to the fact that the login was located as a button, and the registration was also located as a link under the login button, and thus, the participants did not notice the registration link. Additionally, since the target group was low-skilled adults, it was observed that their digital skill level for online activities was quite low.

After login to the system, the demographic form is presented to the participants, and other operations cannot be started without filling all the fields. If there are fields left blank, the save button should be pressed after filling them. Therefore, another common problem was trying to complete the task without filling all the demographic form fields. The participants spent a lot of time in this task until they realized this situation. It may be stated that this was due to the long demographic form and the carelessness of the participants.

In addition to these, failure to find the institution information selection menu and failure to find the test list in the menu were among the important problems. When the screen recordings and observation data were examined, it was determined that the participants clicked on different places on the web page and randomly found the menu after a few tries. However, the menu icon was a standard icon used in digital systems. For this reason, it is thought that these problems were experienced due to the very low digital skill level of the target group.

It was observed that some participants could not use the drag and drop pointer in the sorting question task. In the simulation questions, it was determined that most of the participants did not understand how to respond to the question. Such that, according to the data in Tables 3 and 4, it was seen that the participants P10 and P14 left the system without answering any of these questions. It may be stated that this was again due to the low digital skill level of the target group.

Furthermore, it was determined that some participants found the exit menu by trying (like other problems regarding not finding the menu), and some of them even dropped out of the system by closing the browser. Similarly, although the logout icon was also a standard icon used in digital systems, the users experienced problems here as well due to their low digital skill levels and the small font size of the question root statement.

5. DISCUSSION AND CONCLUSION

In this study, it was aimed to evaluate the usability of OSAT developed to evaluate the literacy, numeracy and digital skills of low-skilled/underqualified adults within the scope of the online skill assessment project. The sample consisted of 14 adults between the ages of 18 and 40 who had graduated from secondary school at most and these formed the target group of the project. In the process of the usability evaluation of OSAT, the user-based method, in which user experiences are employed (Dillon, 2001), and the usability framework determined by ISO 9241-11 (1998) were used. An observation form and a satisfaction questionnaire developed by the researchers were used as the data collection tools. The effectiveness, efficiency and satisfaction attributes of OSAT were evaluated as usability measures. Depending on this, in the usability evaluation of OSAT, according to the data obtained from the screen recordings, effectiveness was considered as the completion rate of the tasks, and efficiency was considered as the completion duration of the tasks. Satisfaction was also considered as values obtained from the satisfaction questionnaire data. The usability test was conducted for the application module, which is the user web interface of OSAT. For this, a total of 20 tasks, including 7 main tasks, were determined. While determining the usability tasks for the application module for which the usability test was performed, to make a comprehensive evaluation, all menu operations and test interfaces of the application module were aimed to be included.

According to the findings, the participants completed most of the tasks at a high rate. Although the mean time to complete each task varied, the participants completed the tasks approximately between 12 and 25 minutes. Gülbahar, Kalelioğlu, & Madran (2008), in their study with 5 students and 5 academicians, evaluated the usability of a higher education web-based assessment and evaluation tool, WebQuest, and concluded that the tasks were completed with a 75% success rate. According to the findings of the satisfaction questionnaire, it is possible to say that the participants were generally satisfied with the use of the system. Karahoca, Karahoca, & Günoğlu (2009) analyzed the usability of a web-based exam automation system they developed in terms of the learnability, controllability, design, and satisfaction factors through a survey and concluded that the system was successful in terms of satisfaction. In another study, Lai, Chen, & Chou (2017) aimed to examine the interface design of a web-based formative assessment system. In the study attended by 28 college students, user experiences were collected through a web assessment usability questionnaire and a web assessment interface survey. So, the results indicated that overall usability scores were high. Similarly, in the usability evaluation of a web-based assessment tool developed for improvement of communication and active listening skills, no technical difficulty was reported, and the participants were satisfied with the system according to the survey data (Cheon & Grant, 2009).

Some usability problems were identified as a result of the observations and the detailed examination of the screen recordings. These problems provided important clues for the updates to be made in the design of the system to make the system more usable before the system is opened to end users. Primarily, one of the situations where the participants experienced the most problems was trying to login to the system without registration. It was determined that this was caused by the participants not noticing the registration link. To solve this problem, the registration link was changed to a button like the login and positioned next to the login button to increase its visibility. Another common problem was trying to complete the task without filling all the demographic form fields. For this situation, an explanation was added at the beginning of the relevant page that all fields are mandatory. Not being able to find the menus was also one of the major problems. Although the most widely used icons were used in the OSAT design, the very low digital skill level of the target group led to these icons not being understood. For this reason, the icons were removed, and the related operations were transformed into text forms such as "menu" and "logout" in order to make it easier for users to understand. In addition to these, it was observed that some participants could not use the drag and drop pointer in the sorting question task. Therefore, the explanation "by dragging up and down" was added to the root statement of the question, and the pointer was made more explicit. In the simulation questions, it was determined that most of the participants did not understand how to answer the questions. It was observed that this problem was due to the small font size of the question root statement and the low digital skill level of the target group. In order to overcome this problem, the font size of the question root statements in the simulation questions was changed, and small hints on how to answer the questions were added.

Although the discussion was limited due to the scarcity of studies on the usability of online assessment systems and the fact that most of them were based on survey data, it is possible to say that task-based usability tests, in which user experiences are employed, are guiding researchers in designing and developing systems. Nowadays, when learning and teaching processes are transferred to online environments, assessment, which is a part of this process, comes to the fore. So, many online assessment tools are used or developed to determine the quality of education. Therefore, it may be stated that more detailed usability studies are needed for these online assessment systems. Moreover, although no matter how the system was designed in line with universal design principles, situations where features such as the prior knowledge and skill level of the target group come to the fore, and as in this study, they oblige the system to be designed specifically for the target users. As a matter of fact, Battal & Çağıltay (2015) emphasized that the user profile that will use the system according to the intended use of the system should also be taken into consideration while designing systems. Finally, for future research, it may be recommended to conduct studies on the usability evaluation of online assessment systems with larger samples, where different methods may be employed, and in-depth findings and indicators may be revealed.

By the way, OSAT, which was developed within the scope of the project supported within the framework of the EaSI Program, sets an example of supporting the dissemination of skills assessment in the implementation of the EU's Upskilling Pathways policy. In this context, the OSAT is expected to serve stakeholders such as Turkish Employment Agencies or Development Agencies, who need individual skills assessment of low-skilled/underqualified adults nationally and/or regionally. Thus, it will be possible to determine the training they need by evaluating the current skill levels of the individuals and to provide the necessary training opportunities, as well as to ensure the sustainability of the developed tool.

Hakem Değerlendirmesi: Dış bağımsız.
Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemiştir.
Yazar Katkıları: Çalışma Konsepti/Tasarımı: M.Y., M.B., Ö.Ö.; Veri Toplama M.Y., F.K., Ö.Ö.; Veri Analizi /Yorumlama: M.Y.; Yazı Taslağı: M.Y., M.B., F.K., Ö.Ö..; İçeriğin Eleştirel İncelemesi: M.Y., M.B., Ö.Ö.; Son Onay ve Sorumluluk: M.Y., M.B., F.K., Ö.Ö.
Finansal Destek: Bu çalışma, Avrupa Birliği'nin İstihdam ve Sosyal Yenilik Programı "EaSI" (2014-2020) kapsamında finansal destek almıştır. Proje Sözleşme Numarası: VS/2019/0127. Daha detaylı bilgi için: http://ec.europa.eu/social/easi

Peer-review: Externally peer-reviewed.
Conflict of Interest: The authors have no conflict of interest to declare.
Author Contributions: Conception/Design of Study- M.Y., M.B., Ö.Ö.; Data Acquisition- M.Y., F.K., Ö.Ö.; Data Analysis/Interpretation- M.Y.; Drafting Manuscript- M.Y., M.B., F.K., Ö.Ö.; Critical Revision of Manuscript- M.Y., M.B., Ö.Ö.; Final Approval and Accountability- M.Y., M.B., F.K., Ö.Ö.
Grant Support: This study was received financial support from the European Union Programme for Employment and Social Innovation "EaSI" (2014-2020), Project Contract Number: VS/2019/0127. For further information please consult: http://ec.europa.eu/social/easi

References/Kaynaklar

Akgül, Y. (2020). Accessibility, usability, quality performance, and readability evaluation of university websites of Turkey: a comparative study of state and private universities. *Universal Access in the Information Society*, 1-14. doi:10.1007/s10209-020-00715-w.

Althobaiti, M. M., & Mayhew, P. (2015, September). Assessing the usability of learning management system: user experience study. In *Second International Conference on E-Learning, E-Education, and Online Training* (pp. 9-18).

Anam, H., Sadiq, M., & Jamil, H. (2020). Development of System Usability Scale (SUS) for the Urdu Language. *International Journal of Computer Science and Information Security, 18*(6), 73-78.

Battal, A., & Çağıltay, K. (2015). Investigation of usage frequency and department on usability ıssues with the Online Examination System (OSSI). *Mersin University Journal of the Faculty of Education, 11*(3), 752-763.

Bayrak, T., & Akcam, B. (2015). Exploring benefits of a web based testing and training tool. *Procedia-Social and Behavioral Sciences, 195*, 1032-1041. Doi:10.1016/j.sbspro.2015.06.146.

Cengiz, E. (2016). Hacettepe Üniversitesi Kütüphaneleri web sitesinin kullanılabilirliğinin eğitimle desteklenmesi ve değerlendirilmesi: Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü öğrencileri üzerine bir çalışma. *Türk Kütüphaneciliği, 30*(1), 84-95.

Chang, C., & Almaghalsah, H. (2020). Usability evaluation of e-government websites: A case study from Taiwan. *International Journal of Data and Network Science, 4*(2), 127-138.

Chao, J. Y., Chao, S. J., Yao, L. Y., & Liu, C. H. (2016). A case study of design and usability evaluation of the collaborative problem solving instructional platform system. *Eurasia Journal of Mathematics, Science and Technology Education, 12*(10), 2647-2655. Doi:10.12973/eurasia.2016.1278a.

Cheon, J., & Grant, M. (2009). Active listening: Web-based assessment tool for communication and active listening skill development. *TechTrends, 53*(6), 24-34. Doi:10.1007/s11528-009-0340-4.

Çağıltay, K. (2016). İnsan bilgisayar etkileşimi ve eğitim teknolojileri. In K. Çağıltay and Y. Göktaş (Eds.) *Öğretim Teknolojilerinin Temelleri: Teoriler Araştırmalar Eğilimler* (pp. 297-314). Pegem Akademi, Ankara.

de Castro Lozano C. et al. (2011). Usable interface design for everyone. In Cipolla Ficarra FV, de Castro Lozano C, Nicol E, Kratky A, Cipolla-Ficarra M (Eds) *Human-Computer Interaction, Tourism and Cultural Heritage* (pp. 157-172). Springer, Berlin, Heidelberg. Doi:10.1007/978-3-642-18348-5_15.

Dillon, A. (2001). The evaluation of software usability. In Karwowski W (Ed) *Encyclopedia of Human Factors and Ergonomics.* Taylor and Francis, London. https://repository.arizona.edu/bitstream/handle/10150/105344/Adusability.pdf?sequence=1&isAllowed=y. Accessed 15 October 2020.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (2004). *Human-computer interaction.* Prentice Hall, England.

Dumit, N. Y. (2012). Diagnostic/formative/summative assessment. https://vb.ckfu.org/attachments/%E3%E1%CA%DE%EC-%C7%E1%CE%D1%ED%CC%ED%E4/303900d1479934616-%E3%D1%CC%DA-%D4%C7%E3%E1-%E1%DF%DD%C7%ED%C7%CA-%C7%E1%DE%D3%E3-%C7%E1%DA%C7%E3-%E6%DF%DD%C7%ED%C7%CA-%C7%E1%CA%CE%D5%D5-%C7%E1%E1%DB%C9-%C7%E1%C7%E4%CC%E1%ED%D2%ED%C9-diagnostic-formative-summative-asst.pdf. Accessed 20 October 2020.

Ersoy, H. (2004). Bir çevrimiçi öğrenim destek sisteminin kullanılabilirlik testi: Planlama, uygulama, değerlendirme. *The Turkish Online Journal of Educational Technology, 3*(1), 75-82.

Gonçalves, R., Rocha, T., Martins, J., Branco, F., & Au-Yong-Oliveira, M. (2018). Evaluation of e-commerce websites accessibility and usability: An e-commerce platform analysis with the inclusion of blind users. *Universal Access in the Information Society, 17*(3), 567-583. Doi:10.1007/s10209-017-0557-5.

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.

Gülbahar, Y., Kalelioğlu, F., & Madran, O. (2008). Usability evaluation of "web macerası" as an instructional and evaluation method. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 41*(2), 209-236.

Güven, Ç. (1987) Okulda öğrenme-öğretme sürecinin denetimi. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi, 2*(1), 271-285.

Harrati, N., Bouchrika, I., Tari, A., & Ladjailia, A. (2016). Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Computers in Human Behavior, 61,* 463-471. Doi:10.1016/j.chb.2016.03.051.

Hasan, L. (2014). The website of the University of Jordan: Usability evaluation. *International Arab Journal of e-Technology, 3*(4), 258-269.

Hasan, L., Morris, A., & Probets, S. (2012). A comparison of usability evaluation methods for evaluating e-commerce websites. *Behaviour & Information Technology, 31*(7), 707-737. Doi:10.1080/0144929X.2011.596996.

Huang, Z., & Benyoucef, M. (2014). Usability and credibility of e-government websites. *Government Information Quarterly, 31*(4), 584-595. Doi:10.1016/j.giq.2014.07.002.

Ismail, A., & Kuppusamy, K. S. (2018). Accessibility of Indian universities' homepages: An exploratory study. *Journal of King Saud University-Computer and Information Sciences, 30*(2), 268-278.

ISO 9241-11 (1998). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. https://pdfs.semanticscholar.org/17e5/3f6fa806ea573e4ee0904fbb12c3a527d5f0.pdf. Accessed 10 October 2020.

Kakasevski, G., Mihajlov, M., Arsenovski, S., & Chungurski, S. (2008, June). Evaluating usability in learning management system Moodle. In *ITI 2008 - 30th International Conference on Information Technology Interfaces* (pp. 613-618).

KAMIS (2018). Kullanılabilirlik ilkeleri. https://kamis.gov.tr/wp-content/uploads/2018/03/Bolum3.pdf. Accessed 12 October 2020.

Karahoca, A., Karahoca, D., & Günoğlu, S. (2009). Web tabanlı sınav otomasyon sisteminin kullanılabilirlik analizi. In *4. Ulusal Yazılım Mühendisliği Sempozyumu* (pp 183-190).

Lai, T. L., Chen, P. Y., & Chou, C. Y. (2017, May). A user experience study of a web-based formative assessment system. In *2017 International Conference on Applied System Innovation* (pp. 899-902).

Lee, S., & Koubek, R. J. (2010). The effects of usability and web design attributes on user preference for e-commerce web sites. *Computers in Industry, 61*(4), 329-341. Doi:10.1016/j.compind.2009.12.004.

Li, F., & Li, Y. (2011). Usability evaluation of e-commerce on B2C websites in China. *Procedia Engineering, 15*, 5299-5304. Doi:10.1016/j.proeng.2011.08.982.

Lin, H. C. K., Hsieh, M. C., Wang, C. H., Sie, Z. Y., & Chang, S. H. (2011). Establishment and usability evaluation of an interactive AR learning system on conservation of fish. *Turkish Online Journal of Educational Technology, 10*(4), 181-187.

Lyzara, R., Purwandari, B., Zulfikar, M. F., Santoso, H. B., & Solichah, I. (2019, January). E-government usability evaluation: Insights from a systematic literature review. In *Proceedings of the 2nd International Conference on Software Engineering and Information Management* (pp 249-253).

Majzub, R. M. (2009). The development of a web based ecological assessment of school readiness (WEBEASR). *Procedia-Social and Behavioral Sciences, 1*(1), 2568-2572. Doi:10.1016/j.sbspro.2009.01.453.

Mentes, S. A., & Turan, A. H. (2012). Assessing the usability of university websites: An empirical study on Namik Kemal University. *Turkish Online Journal of Educational Technology, 11*(3), 61-69.

Newell, A. F., & Gregor, P. (2002). Design for older and disabled people–where do we go from here?. *Universal Access in the Information Society, 2*(1), 3-7. Doi:10.1007/s10209-002-0031-9.

Nielsen, J. (1994). *Usability engineering.* Morgan Kaufmann.

Ogolla, J. A. (2011). Usability evaluation: Tasks susceptible to concurrent think-aloud protocol. Dissertation, Linköping University.

Pant, A. (2015). Usability evaluation of an academic library website: Experience with the Central Science Library, University of Delhi. *The Electronic Library, 33*(5), 896-915. Doi:10.1108/EL-04-2014-0067.

Paul, S., & Das, S. (2019). Accessibility and usability analysis of Indian e-government websites. *Universal Access in the Information Society, 19,* 949–957. Doi:10.1007/s10209-019-00704-8.

Pensabe-Rodriguez, A., Lopez-Dominguez, E., Hernandez-Velazquez, Y., Dominguez-Isidro, S., & De-la-Calleja, J. (2020). Context-aware mobile learning system: Usability assessment based on a field study. *Telematics and Informatics, 48*, 101346. Doi: 10.1016/j.tele.2020.101346.

Phongphaew, N., & Jiamsanguanwong, A. (2017, July). Usability evaluation on learning management system. In *International Conference on Applied Human Factors and Ergonomics* (pp 39-48).

Sagar, K., & Saha, A. (2017). A systematic review of software usability studies. *International Journal of Information Technology.* Doi:10.1007/s41870-017-0048-1.

Seffah, A., Donyaee, M., Kline, R. B., & Padda, H. K. (2006). Usability measurement and metrics: A consolidated model. *Software Quality Journal, 14*(2), 159-178. Doi:10.1007/s11219-006-7600-8.

Senemoğlu, N. (2013) *Gelişim, öğrenme ve öğretim. Kuramdan uygulamaya* (23.baskı). Yargı Yayınevi, Ankara.

Shi, L., Al Qudah, D., Qaffas, A., & Cristea, A. I. (2013, June). Topolor: a social personalized adaptive e-learning system. In *International Conference on User Modeling, Adaptation, and Personalization* (pp 338-340).

Story, M. F., Mueller, J. L., & Mace, R. L. (1998). The universal design file: Designing for people of all ages and abilities. https://files.eric.ed.gov/fulltext/ED460554.pdf. Accessed 15 October 2020.

Şahin, İ., Hebebci, M. T., & Çelik, İ. (2014, May). Çevrimiçi öğrenme ortamlarında kullanılabilirlik: Bir literatür tarama çalışması [Usability in the online learning environments: A literature review study]. In *International Conference on Education in Mathematics, Science & Technology* (pp 1354-1362).

TÜİK (2019). Sorgulama ekranı. https://biruni.tuik.gov.tr/medas/. Accessed 11 October 2020.

Vairamuthu, S., & Anouncia, S. M. (2016). Reconnoitering students' satisfaction of an online based assessment system to improve usability using PSO: An examination into a problem solving and programming course. *Engineering, Technology & Applied Science Research, 6*(5), 1207-1211. Doi:10.48084/etasr.800.

Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology, 22*(5), 339-351. Doi: 10.1080/0044929031000.

Ventayen, R. J. M., Estira, K. L. A., De Guzman, M. J., Cabaluna, C. M., & Espinosa, N. N. (2018). Usability evaluation of google classroom: Basis for the adaptation of gsuite e-learning platform. *Asia Pacific Journal of Education, Arts and Sciences, 5*(1), 47-51.

Yıldırım, A. & Şimşek, H. (2013). Sosyal bilimlerde nitel araştırma yöntemleri (9.baskı). Seçkin Yayıncılık, Ankara.

APPENDIX

Satisfaction Questionnaire

Item No		Items	(1) Strongly disagree	(2) Disagree	(3) Neutral	(4) Agree	(5) Strongly agree
1	Design	The font and font size in the system make it easy to read.					
2		The harmony of the background color and text color in the system makes it easy to read.					
3		The images in the system and on questions are comprehensible.					
4		The colors of the menus in the system are appropriate.					
5		The colors of the form elements (such as option buttons, input boxes, checkboxes) are appropriate.					
6		The colors of the navigation elements (such as buttons, scroll bars, drag-and-drop markers) are appropriate.					
7	Navigation	The placement of the pages in the system on the screen is appropriate.					
8		The placement of the menus in the system on the screen is appropriate.					
9		The placement of the form elements (such as option buttons, input boxes, checkboxes) in the system on the screen is appropriate.					
10		The placement of the navigation elements (such as buttons, scroll bars, drag-and-drop markers) in the system on the screen is appropriate.					
11		The clickable items (links) in the system are comprehensible.					
12		The content in the system is easy to follow.					
13	Ease of use	Generally, the system is easy to use.					
14		The help menu is not needed in order to understand the usage of the system					
15		The instructions and explanations in the system are sufficient.					
16		All pages and links in the system work smoothly.					