

Research Article

**Designing a Usability Assessment Process for Adaptive Intelligent Tutoring Systems:
A Case Study¹**

Ali Kürşat Erümit², İsmail Çetin³, Mehmet Kokoç⁴, Temel Kösa⁵,
Vasıf Nabiyev⁶, Emine Selin Aygün⁷

Abstract

This study aims at examining the usability evaluation process of ArtiBos, an adaptive intelligent tutoring system specifically designed for teaching of problem solving, and proposing usability evaluation recommendations accordingly. The study group consists of 90 students selected from 3 different secondary education institutions and 30 experts from the fields of Computer Education and Instructional Technology and Mathematics Teaching. The usability of ArtiBos is evaluated according to the criteria of effectiveness, efficiency, and satisfaction included in the definition of usability by the International Standards Organization (ISO). In this study, quantitative and qualitative data were used for triangulation. Study data included users' screenshots during performance of the instructed tasks in ArtiBos, sound recordings taken from the think aloud technique, interviews with students, and opinions of experts. The study elaborates the procedure of usability tests of ArtiBos and makes suggestions for usability evaluation processes of similar systems. The results are expected to guide researchers planning to carry out similar systems' usability studies.

Keywords: *Usability, intelligent tutoring systems, expert based evaluation, user based evaluation*

¹ This research is supported by The Scientific and Technological Research Council of Turkey (Project Number: TÜBİTAK-215K029)

² Asst.Prof.Dr., Trabzon University, Fatih Faculty of Education, Department of Computer Education and Instructional Technology, kursaterumit@gmail.com, <https://orcid.org/0000-0003-4910-4989>

³ Res.Asst., Ondokuz Mayıs University, Institut of Educational Sciences, ismail.cetin@omu.edu.tr, <https://orcid.org/0000-0002-7865-6080>

⁴ Res.Asst.Dr., Trabzon University, Fatih Faculty of Education, Department of Computer Education and Instructional Technology, kokoc@trabzon.edu.tr, <https://orcid.org/0000-0002-1347-8033>

⁵ Asst.Prof.Dr., Trabzon University, Fatih Faculty of Education, Department of Mathematics and Science Education, temelkosa@ktu.edu.tr, <https://orcid.org/0000-0002-4302-1018>

⁶ Prof.Dr., Karadeniz Technical University, Faculty of Engineering, Department of Computer Engineering, vasif@ktu.edu.tr, <https://orcid.org/0000-0003-0314-8134>

⁷ Graduate Student, Trabzon University, Department of Computer Education and Instructional Technology, 346030@trabzon.edu.tr, <https://orcid.org/0000-0003-0759-8095>

Uyarlanabilir Zeki Öğretim Sistemleri için Kullanılabilirlik Değerlendirmesi Süreci Tasarım Önerisi: Durum Çalışması

Öz

Bu çalışmanın amacı, problem çözme öğretimi için tasarlanmış bir uyarlanabilir zeki öğretim sistemi olan ArtiBos'un kullanılabilirlik değerlendirmesi sürecini incelemek ve benzer sistemlerin kullanılabilirlik değerlendirmesi süreçlerine yönelik öneriler sunmaktır. Çalışmaya, 3 farklı ortaöğretim kurumundan seçilen 90 öğrenci ile Bilgisayar ve Öğretim Teknolojileri Eğitimi-Matematik Eğitimi alanlarından 30 uzman katılmıştır. ArtiBos'un kullanılabilirliği, Uluslararası Standartlar Enstitüsü (International Standards Organization – ISO) tarafından yapılan kullanılabilirlik tanımında yer alan etkililik, verimlilik ve memnuniyet kriterlerine göre değerlendirilmiştir. Çalışmada veri çeşitlemesi sağlanarak nicel ve nitel veriler birlikte kullanılmıştır. Kullanıcıların ArtiBos'da istenen görevleri gerçekleştirme süreçlerindeki ekran görüntüleri, sesli düşünme tekniği (think aloud) sırasında alınan ses kayıtları, öğrencilerle yapılan görüşmeler ve uzmanların görüşleriyle veriler elde edilmiştir. Çalışmada, ArtiBos'un kullanılabilirlik testleri süreci detaylı bir şekilde açıklanmış ve kullanılabilirlik değerlendirme süreçlerine yönelik öneriler sunulmuştur. Elde edilen sonuçların benzer sistemlerin kullanılabilirlik çalışmasını yapmayı planlayan araştırmacılar için yol gösterici olması hedeflenmektedir.

Anahtar Sözcükler: *Kullanılabilirlik, zeki öğretim sistemleri, uzman temelli değerlendirme, deneysel değerlendirme*

Introduction

Today computers are being used in every area of life by people from all walks of life. For this situation, a big share is owed to software interfaces. Until the 1970s, computers could only be used by experts for military, academic, and commercial experts chiefly because computers were designed for specific tasks and it was experts only who had a command of the tasks required for those tasks. However, the situation began to change upon introduction of Simula and SmallTalk in 1967 and 1971, respectively, the first object-oriented programming languages. As such programming languages and derivatives were put into use, software products which did not previously have an interface started to be replaced by versions that are easier to use and can be understood by all. Interface design has become quite advanced and become one of the most important elements of software since then (Reilly, 2003). Norman (1988) points out that the most important feature of a software program is a convenient design and minimized probability of errors. This might be possible if the interface is evaluated together with stakeholders from each segment of users and the design is made usable for those people. Myers and Rosson (1992) found out that the interface accounts for 48% of satisfaction with a program developed in their study. This result shows that the interface design is as important as the content in a software.

The field which studies the relationship between interface design of computer programs and users is called Human-Computer Interaction (HCI). The discipline works ensure effective use of technology and to create, upgrade, and evaluate technology to facilitate human life (Özdemir, Atasoy, & Somyürek, 2007). HCI is regarded interdisciplinary because it is borrows from a variety of disciplines (Carroll, 2003; Çağıltay, 2011; Lazar et al., 2010). Of the procedures referred to in definition of HCI, studies on usability at design stage come to the forefront. The concept of usability, which is also known as easy to use, user-friendly, and transparent to the user, is difficult to measure and has a complex structure; therefore, it currently lacks a scientific definition agreed upon by all sectors (Hertzum, Hansen, & Andersen, 2009). Nonetheless, frequent everyday brand slogans highlighting ease of use can be examples of semantic definition; such as “Easy to connect”, “programmed easily”, and “easy-open bag”.

According to Nielsen (2012), a veteran of studies on usability, usability refers to a quality indicator that measures how easily the interfaces are used, and to methods developed to

increase the ease of use during design. Another important standard related to usability comes from the definition by ISO, which targets two aspects as quality of use and software quality. The former means the capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified contexts of use (ISO 9241-11, 1998). Software quality refers to the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions (ISO/IEC 9126-1, 2001). In other words, usability of a developed system must meet the criterion of satisfaction besides effectiveness and efficiency.

Usability tests to assess the usability of a product consist of 4 different approaches. These are expert-based approach (heuristic evaluation), approach based on design guidelines, experimental approach (user test), and model-based approach (Çağiltay, 2011).

In addition to approaches, there are also test types: in-process tests and end-of-process tests. In in-process tests, the product evaluation starts as soon as the product is designed, whereas the evaluation takes place after the product is designed in the other type. In-process tests are quite important to predetermine design problems in the product and to discover aspects that could be hardly modified later. Still, those tests are not fault-proof. End-of-process evaluation must be also performed as a pre-caution. Considering pros and cons, it seems more robust as an approach to use both types together (Çağiltay, 2011).

Usability studies have generally been done for commercial and academic purposes on websites. Examples of studies on usability of public web sites (Durmuş & Çağiltay, 2012; Arsoy, Kalıpsız, & Öztürk, 2013; Yavuz, Çınar, & Çağiltay, 2016), university web pages (Cevher, 2015; Çebi, Durucu, & Kayhan, 2013), university libraries (Cockrell & Jayne, 2002; Cengiz, 2016; Iqbal & Ullah, 2016), and e-commerce sites (Lee & Koubek, 2010; Zviran, Glezer, & Avni, 2006), and academic systems Özdemir, Atasoy, & Somyürek, 2007) and academic databases (Çetin & Şendurur, 2016). Besides web pages; usability studies exist about educational environments (Bayram & Yeni, 2011; Beymer, Orton, & Russell, 2007; Can, Atalay, & Eraslan, 2017; Pala, Arslan, & Özdiñç, 2017; Erdoğan & Şahin, 2018) and mobile applications (Dönmez, Yaman, Şahin, & Yurdakul, 2016; Oyibo, Ali, & Vassileva, 2016). There are also usability studies with different purposes such as investigating the effect of the ads appearing along with search results on the Internet on users (Buscher, Dumais, & Cutrell, 2010).

The main goal of usability studies in education is to make educational softwares more student-friendly with the aim of easing the cognitive burden that may arise from the design of the material, thus increasing the efficiency and efficiency of the educational environments through increased focus on the content. Today, a number of innovations are introduced to improve the effectiveness and efficiency of digital learning environments. Blending of new technologies and approaches in the design of user-centered learning environments is gaining importance. A number of new technologies are on the ground to provide a richer learning experience for students. They include distance learning technologies for time and space-independent learning, digital game-based teaching to ensure lasting and effective learning, individualized teaching to provide customized training, adaptive teaching, and intelligent tutoring systems.

Intelligent tutoring systems (ITS) are computerized teaching environments based on artificial intelligence technology to support learning and teaching processes (Bernacki et al., 2014). These systems give support to the student throughout the teaching process by simulating the instructor's teaching style (Magnisalis, Demetriadis, & Karakostas, 2011). ITSs are comprised of 4 modules as user interface module, student module, teaching module, and information module (Koedinger & Corbet, 2006; Victorio-Meza, Mejialavalle, & Ortiz, 2014; Vos, 1995). The main task of ITSs is to evaluate students' knowledge acquisition throughout the process. Evaluation is necessary to adapt learning materials and activities to student data (Ramirez- Norigea, Juarez-Ramirez, & Martinez-Ramirez, 2017). Adaptation refers to the task of designing content or navigation suggestions according to the data of individual user characteristics in view of the objectives of the training (Radenkovic, et al., 2011). Adaptive systems aim to tailor learning environments for learners (Reniers & Dreher, 2009). Such systems are rapidly becoming widespread nowadays and software and technology products developed in this field are increasing (Tuna & Öztürk, 2015).

Adaptive tutoring systems discover the needs and wishes of the student and customize the system in this direction, while intelligent systems simulate the style of the real teacher and provide support to the individual learner during the teaching period (Magnisalis, Demetriadis, & Karakostas, 2011). In broad terms, adaptive intelligent tutoring systems (AITSs) are systems based on artificial intelligence technology, which can be used to support learning and teaching processes, by saving a student's preferences during the use of the software, which

can determine the student's competencies and deficiencies, and then use this information to adapt itself and guide the student (Bernacki et al., 2014).

Aim of the Study

In the literature, there are abundant interface design usability studies concerning computer-aided materials, while this number is quite low for adaptive intelligent teaching systems. In general, research focuses on architectural design rather than interface design (Chughtai, Zhang, & Craig, 2015). It is thought that recommendations to follow the evaluation of the system interfaces and the results will bring about design of more efficient and efficient learning environments. Therefore, this study intends to assess the design and usability of ArtiBos, an adaptive intelligent teaching system, was made. The aim of the study is to provide guidance for usability evaluation of systems similar to ArtiBos while performing the usability tests of ArtiBos, and to make recommendations accordingly. For this purpose, the study seeks answers the following research questions.

1. How can the process of usability evaluation be planned for AITSSs?
 - 1.1. What has been done in the usability evaluation processes of the AITSSs?
 - 1.2. What has been done in the usability evaluation process of ArtiBos?
2. How is ArtiBos's usability?

Method

In this study, findings on usability studies related to adaptive tutoring systems and intelligent tutoring systems were analyzed by using document analysis method and then the steps to be taken in ArtiBos's usability study was determined. After that, according to these steps, a usability evaluation of ArtiBos was carried out with both expert and student views. In study,

- Document analysis was performed on databases with specific keywords in order to determine the steps to be followed during the usability studies.
- Experimental approach was applied for the usability test by conducting user tests and interviews with students. Moreover, data collected from students by means of think aloud were recorded and screen recording was made;

- Expert-based approach was utilized where experts were interviewed about intuitional analysis and system design evaluation.

In order to obtain detailed information in this process, various tools of data collection were used such as article review form for document analysis, usability test tasks for the usability test with the students, a semi-structured interview form for the system evaluation interview with the students, Nielsen's (1994) heuristic rubric for the heuristic analysis with the experts, and system design evaluation form for the design evaluation interview with the experts.

Considering the procedures, data collection tools, and the analysis methods in this study; it can be said to be implemented with embedded design as a type of mixed research. Embedded design is a mixed-method pattern which brings together quantitative and qualitative data concurrently or sequentially (Creswell, 2008). Basically qualitative methods are used in the study; yet, quantitative data were also collected for variety in accordance with the embedded mixed research design.

Data Collection and Analysis

In the process of examining the studies, articles published after 2000 were examined in Google Scholar, Eric, Sciencedirect and National Thesis Center databases by using “intelligent tutoring systems”, “adaptive systems” and “usability” keywords. The title, summary and keywords of the articles were examined in order to determine whether the studies obtained during the search process will be evaluated within the scope of the research. When the information in these sections is not sufficient, other sections of the articles are examined. 13 articles providing the criteria for preliminary examination were selected for further analysis and included in the study. These articles were analyzed and coded according to the criteria of the purpose, type, study group, data collection tool and usability approach. The selection and coding of the articles was carried out by two researchers. The results of the analysis were presented as a table. The collection and analysis of data on the usability process were described under Data Collection and Data Analysis.

Population and Participants

The population in the document analysis process of study are article published in refereed journals about adaptive tutoring systems, intelligent tutoring systems and adaptive intelligent tutoring systems. The participants in the usability evaluation process is explained under selection of participants.

The Media Used – ArtiBos as an Adaptive Intelligent Tutoring System

Developed as an adaptive intelligent tutoring system, ArtiBos consists of 7 main modules and sub-modules placed under them. There are 7 modules which can be accessed by users by logging in ARTIBOS home screen with their user name and passwords. They are Lecture Module, Problem-Building and Editing Module, Problem Solving Module, Problem Asking Module, Users Account Modul, Problem Level and Score Determination Module and Adaptive Module.

Lecture module is designed in a way to include the basic concepts related to the topic to be taught. The design of Problem-Building and Editing Module allows students to configure problems on their own. Problem Solving Module is the module that allows solving the problems to be built by students and then to be stored on a common server to be provided by the students and their peers. In Problem Asking Module where problems are sent, created problems can be asked to all other online users or to one specific competitor as a challenge or duel. In Users Account Module, students edit their personal information and view questions they added and their score. Problem Level and Score Determination Module is used for scoring in the system. And in Adaptive Module, Adaptation is performed as adaptation of the content for the scene, object pools and applicable difficulty level, taking into account the student's success level. Figure 1 shows the general view of ArtiBos.

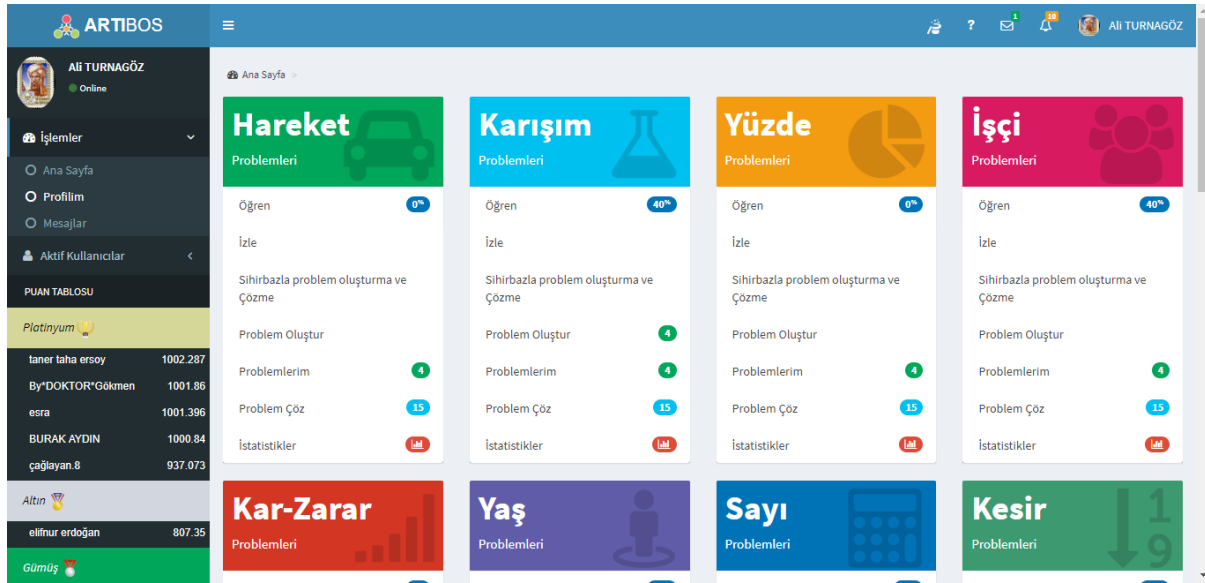


Figure 1. General view of ArtiBos

Scenes to be created in the system are recorded as video and text. When the student wants to ask the problem, the problem is sent to the other side as an animating object (video) together with the problem text. As the student wishes to address the problem; such data are accompanied by various paths to solve the problem, the operation steps required for the solution and the result of the problem are uploaded to the Web server. ArtiBos features an audio and visual educational calendar in order to follow the students' steps and give the appropriate feedback during the problem formation activities. In order to carry out an inclusive evaluation while determining the usability test tasks, efforts were made to include all of the user-interactive modules of ArtiBos.

Results

In this study, firstly, the usability evaluation processes of ITSs and AITSs were examined. Then, it was explained that what is done in the usability evaluation process of ArtiBos. Finally, an improved usability evaluation process was applied.

Usability Evaluation Processes of ATs and ITSs

Although there are few direct studies on interface design of ITSs and adaptive systems, examples are available that examine the effectiveness and efficiency of the overall structure and content of the system (Roscoe, et al., 2014; Dexheimer, et al., 2017), investigate users'

thoughts and satisfaction with the system (Lin, Wu & Hsueh, 2014; Sanchez, et al., 2014; Verkuyl, et al., 2016), and offer models for a more useful system design (Arevalillo-Herrarez, et al., 2017; Ramirez-Norigea, Juarez-Ramirez & Martinez-Ramirez, 2017).

Below Table 1 displays the usability studies employing various methods for the evaluation of interface and contents both during and after design of intelligent tutoring systems and adaptive teaching systems.

Table 1

Usability Studies on Adaptive and Intelligent Tutoring Systems

Study	Aim	Type	Research and Characteristics		
			Work Group	Data Collection Instrument	Usability Approach
Writing Pal Affective Tutoring System (Roscoe, et al., 2014)	Mobile application prepared to improve students' writing competence	Gamified Intelligent Mobile Application	Performance monitoring with 141 tenth graders	Built-in user test to use the entire system	Experimental approach
			Interviews with 2 experts (English language teachers)	Semi-structured interview form	Expert-based approach
Virtual gaming to develop students' pediatric nursing skills: A usability test (Verkuyt, et al., 2016)	Prepare a better-quality and alternative learning environment for nursing students	Game-based Intelligent Learning Environment	Evaluation of the system with 2 experts	Nielsen (1994) Experts' intuitional evaluation test	Expert-based approach
			6 students of nursing and 5 nurses	Data recorded by voice thinking technique, 16-item usability scale	Experimental approach
Evaluation module based on Bayesian networks to Intelligent Tutoring Systems (Ramirez-Norigea, Juarez-Ramirez, & Martinez-Ramirez, 2017)	Add an evaluation module to an existing intelligent teaching system to perform usability work	Adaptive Intelligent Tutoring System	62 students	Student performance monitoring and student reports	Experimental approach
Affective Tutoring System (Lin, Wu, & Hsueh, 2014)	System prepared for teaching of accounting	Intelligent Tutoring System	80 students	Interviews with students (80 students), student observation form (80 students) and scale of usability for students (45 students)	Experimental approach

Designing a Usability Assessment Process for Adaptive Intelligent Tutoring Systems: A Case Study

Study	Aim	Type	Work Group	Research and Characteristics	
				Data Collection Instrument	Usability Approach
Self-Monitoring Activity-Restriction and Relaxation Treatment (SMART application) (Dexheimer, et al., 2017)	Intelligent Mobile Application for Patients Treated due to Brain Injuries	Intelligent Mobile Application	4 children (aged 11 to 18) and 4 parents (average age of 41)	Think aloud data recorded in 60-minute free use and 5-likert scale of usability	Experimental approach
Ax2ELS (Adaptable-Adaptive English Learning Support) – Teaching of Foreign Language (Sezer, 2011)	Improve foreign language learning	Adaptive Intelligent Tutoring System	Experts of foreign languages (indefinite number of experts)	Field experts evaluating the usefulness and effectiveness of the system	Expert-based approach
Episodic Learner Model Adaptive Remote Tutor (ELM-ART) (Weber & Brusilovsky, 2001)	Present all learning materials as an online interactive textbook	Adaptive E-book Application	Primary school pupils	Questionnaire given to students in order to learn about the effectiveness and usefulness of the application	Experimental approach
INSPIRE (Papanikolaou & Grigoriadou, 2003)	Propose a model to meet students' needs, to evaluate their learning style preferences, and to provide their interactions	Intelligent Tutoring System	66 students	Participants divided into two experimental groups of 33 each in order to check usefulness of the application based on students' feedback	Experimental approach

Study	Aim	Type	Research and Characteristics		
			Work Group	Data Collection Instrument	Usability Approach
Style-OLM (Open Learner Modelling) (Dimitrova, 2003)	Test and evaluate the cognitive capacity of students in order to design and evaluate a tutoring system for them	Game-based Adaptive Intelligent Instruction System	7 graduate students in computer department	Usability questionnaire	Experimental approach
AES-CS (Adaptive Educational System Based on Cognitive Styles) (Triantafillou, Pomportsis, & Georgiadou, 2003)	Design a teaching system with elements for cognitive styles to improve student interactions and learning outcomes	Intelligent Tutoring System	5 experts	Semi-structured interview form	Expert-based approach
			10, 4 th graders	Semi-structured interview form and usability questionnaire	Experimental approach
PEL-IRT (Personalized E-Learning System Based on Item Response Theory) (Chen, Lee, & Chen, 2004)	Design an instructional system that offers course materials and increases individual learning skills	Adaptive Intelligent Tutoring System	210 graduate students	Usability questionnaire	Experimental approach
Personalized Intelligent Tutoring System (PITS) (Chen & Duh, 2008)	Develop a customizable web-based course system application on the classical test theory	Adaptive Tutoring System	High school students (in an indefinite number)	5-point Likert type scale of usability	Experimental approach
Developing an Adaptive Web-Based Intelligent Tutoring System using Mastery Learning Technique (Kularbphetong, Kedsiribut, & Roonrakwit, 2015)	Develop an adaptive web-based intelligent tutoring system using mastery learning technique	Adaptive Web Based Intelligent Tutoring System	67 university students	Usability questionnaire	Experimental approach

When Table 1 is examined, it is seen that the experimental approach is predominantly preferred in evaluation of the usability of adaptive intelligent tutoring systems and the approach is often applied with scales and questionnaires on usability. Some other studies are carried out with think aloud and performance monitoring techniques under experimental research. Apart from the experimental approach, in some studies, expert-based approach is preferred for the evaluation of the systems. In some of these studies, expert-based approach accompanies experimental approach, while it is used as the only method in some others.

The Usability Evaluation Process of ArtiBos

The procedures carried out in ArtiBos’s usability evaluation process and similar procedures in the literature regarding the usability evaluation of the AITs are given in Table 2 and Table 3.

Table 2

Usability Evaluation Approach Applied in ArtiBos

ArtiBos	Literature
Experimental Approach	Roscoe, et al., 2014; Verkuyl, et al., 2016; Ramirez-Norigea, Juarez-Ramirez, & Martinez-Ramirez, 2017; Lin, Wu, & Hsueh, 2014; Dexheimer, et al., 2017; Weber & Brusilovsky, 2001; Papanikolaou & Grigoriadou, 2003; Dimitrova, 2003; Triantafillou, Pomportsis, & Georgiadou, 2003; Chen, Lee, & Chen, 2004; Chen & Duh, 2008; Kularbphettong, Kedsiribut, & Roonrakwit, 2015.
Expert-Based Approach	Roscoe, et al., 2014; Verkuyl, et al., 2016; Sezer, 2011; Triantafillou, Pomportsis, & Georgiadou, 2003.

Table 3

Usability Evaluation Tests Used in ArtiBos

ArtiBos	Literature
User Test	Roscoe, et al., 2014; Ramirez-Norigea, Juarez-Ramirez, & Martinez-Ramirez, 2017; Lin, Wu & Hsueh, 2014; Dexheimer, et al., 2017; Weber & Brusilovsky, 2001; Papanikolaou & Grigoriadou, 2003; Dimitrova, 2003; Chen, Lee, & Chen, 2004; Chen & Duh, 2008; Kularbphettong, Kedsiribut, & Roonrakwit, 2015.
Think Aloud Technique	Verkuyl, et al., 2016; Dexheimer, et al., 2017.
Student Interview	Triantafillou, Pomportsis, & Georgiadou, 2003.
Expert Interview	Roscoe, et al., 2014; Sezer, 2011; Triantafillou, Pomportsis, & Georgiadou, 2003.
Heuristic Analysis	Verkuyl, et al., 2016.

According to table 2 and table 3, in this study, we used experimental approach and expert-based approach as a usability evaluation approach; we used user test, think aloud technique, student interview, expert interview and heuristic analysis as a usability evaluation test. Besides, tables show that in most studies, experimental approach and user test were used. In a few studies, expert-based approach, think aloud technique, student interview, expert interview and heuristic analysis were used. These results show that, most of usability approaches and usability tests were used in this study.

Figure 2 shows that what is done for ArtiBos’s usability evaluation process.

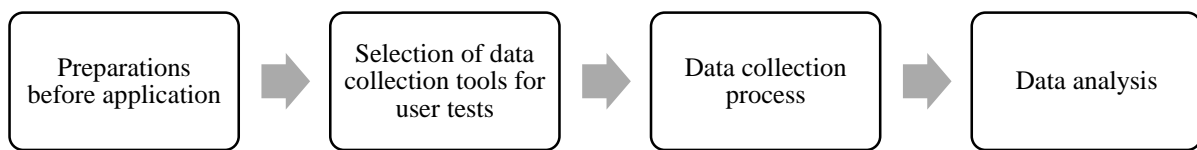


Figure 2. Usability evaluation process

Preparations Before Application

Figure 3 shows that pre-application procedures related to the evaluation process with the students.

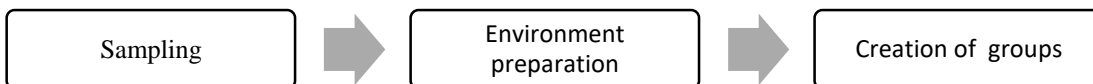


Figure 3. Preparations with students before application

Selection of Participants

In order to encourage focus on the usability rather than the content on ArtiBos, students in the 10th grade were preferred. That is, we preferred purposive sampling technique. In this scope, 90 students attending the 10th grade were selected from 3 Anatolian High Schools (30 students from each of Yavuz Sultan Selim Anatolian High School (YSSAHS), Fatih Sultan Mehmet Anatolian High School (FSMAHS), and Cumhuriyet Anatolian High School (CAHS) school) who had been taught types of problems during the 9th grade (having knowledge of “Applications on equations and inequalities”), and they were assigned tasks on

ArtiBos. Later, the design of the interface of the system was evaluated with 30 design experts comprising of 8 faculty members from the department of Computer Education and Instructional Technologies (CEIT) and 16 PhD students and 6 graduate students majoring in the same department. The reason why all of the participants are not faculty members is that the number of faculty members of the CEIT department is low and the data collection process takes a long time. For this reason easily accessible sampling method was preferred. When choosing experts, their professional experience has been considered. Two of the researchers in the data collection process are graduate students in the Department of Computer Education and Instructional Technologies and a person is a PhD student in the same department.

Environment Preparation

In order to prevent interruption by noise, the libraries of the schools were made available to the researchers during data collection. The researchers also took laptops to the schools so that the students could perform their tasks. All of the laptops are equipped with identical hardware and software features that run at an equal speed. In addition, an application called Camstudio 2.7.4 was installed for screen recording.

Creation of Groups

The students took part in tests as groups of maximum 10 people in each school. The teachers were asked to appointing students from different levels of success as much as possible. The school administration was asked about the students' idle classes, the computer lesson teachers were consulted, and the best time was arranged for the tests.

Selection of Data Collection Tools for User Tests

The procedures for determining the data collection tools for user tests are shown in Figure 4.

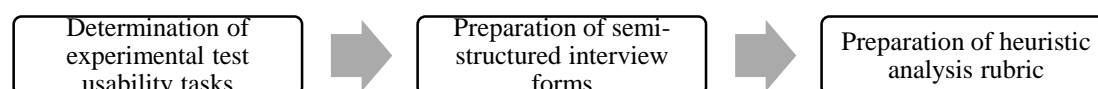


Figure 4. Selection of data collection tools for user tests

Determination of Experimental Test Usability Tasks

In the preparation of usability test tasks, students were able to use general system usage and all menus in detail. The selected tasks were checked and assessed by 4 design experts who are knowledgeable about the system. Then the tasks were updated and finalized. The final version of the tasks was completed by 3 different people who can use a computer fairly and have a command of the system. The average completion duration was prepared as a model for use in the analysis.

Preparation of Semi-Structured Interview Forms

For the students and design experts to evaluate the usability of the system, the semi-structured interview form was prepared by the researchers by paying particular attention to include items covering the entire interface and content of the system.

Preparation of Heuristic Analysis Rubric

As for the rubric targeting the experts, it was drawn up as a 5-point Likert type scale based on the 10 heuristics of Nielsen for usability.

Data Collection Process

In this study, the usability process of the designed system was carried out in 3 distinct schools. Throughout the implementation process, diaries were kept by the researchers to give information about the environment, participants, tests, and challenges. The usability study was completed with both in-process and end-of-process tests. As soon as the system design was launched, it was started to continuously examine and revise its interface by the design experts. The process was carried out with expert-based approach only because the product was not fully ready at that time. However, both of expert-based and experimental approaches were applied when the product was ready. The latter approach was preferred during the testing by users, which consists of tasks for active use of all menus in the system. During the test, the users were asked for their opinions via screen recording and voice thinking. This approach provides more valid and valuable data as it allows access to data from primary users

(Çetin & Şendurur, 2016). As a part of expert-based approach, a rubric was prepared by borrowing items in Nielsen's (2010) heuristics as one of the most important heuristics and the rubric was filled by the design experts for evaluating the system. The number of experts was kept at a maximum level as a precaution for validity of the results. Apart from this, design guidelines and model-based approach were not preferred because design guide approach is often taken by large companies for evaluating their own products and results of this approach are not considered objective enough. The other approach, model-based, seems to be particularly useful for step-by-step systems. Also, practically it is difficult to find an experienced expert in models.

In this scope, experimental approach was taken first and the system was tested by 90 students attending three Anatolian high schools as end users. The study was implemented by recording screen shots and think aloud notes of users while carrying out the assigned tasks. It lasted one week to complete data collection in each school, totaling to 3 weeks. Inside the schools, it took approximately 80 hours to set up computers and media, introduce the system, motivate the students, and to answer the usability test by each group of 10 students.

Screen shots were collected from 30 students using Camstudio 2.7.4, and think aloud data were collected from 90 students by 3 different observers by means of keeping detailed notes. After the tasks, semi-structured interviews were conducted with all of the participants in order to get extensive views of the students about the designs when the students were asked to evaluate each module separately. During the test, the students were told to use the modules of ArtiBos to perform a number of tasks such as sending messages, updating information, creating problems, and solving problems. Photo 1 shows some images taken during the application of the test.



Photo 1. Implementation of the Usability Test

90 Students selected from 3 different high schools were regrouped in groups of maximum 10 people. The test was carried out in the library in each school as the quietest place. All of the participant students were provided with computers with equal hardware and software features. During the data collection, 3 field experts monitored the students in groups of two or three to help them as they needed and record the students' views as they were thinking loudly. At the same time, the researchers noted down the students' views in case the voice recordings were not clear enough. Approximately 80 minutes were allocated for collecting data from each group of students. In order to properly organize the days and times for data collection, the relevant school administration was contacted earlier and appointments were made when the participant students had idle classes or when the teachers planned to teach nothing new as a part of the ordinary classes.

For expert-based assessment, 30 experts completed the rubric covering Nielsen's (1994) heuristics. After filling these rubrics, the experts completed a design evaluation form consisting of 30 questions face to face. While doing the latter, the researchers inquired the

design evaluation questions on the ArtiBos screen and noted the answers. The rubric was designed for scoring between 1 and 5.

The faculty members, research assistants, and postgraduate students in Computer Education and Instructional Technologies were appointed on a volunteer basis. First of all, the experts were trained about the system. Then they answered some open-ended questions after receiving guidance on all menus of the system. Finally, they filled in rubric built on Nielsen's heuristics. The experts carried out the evaluation individually by using their personal computers in an average duration of one hour.

Data Analysis

The usability of the system was judged against the effectiveness, efficiency and satisfaction criteria in the definition by ISO (9241-11, 1998). The effectiveness of the system was checked through the completion status of the tasks, efficiency was judged against the completion time of the tasks, and user satisfaction was evaluated by interpreting checking the notes taken during aloud thinking and the semi-structured interviews with the students following the test. Finally, interviews with experts and rubric results were evaluated for system design.

In order to examine the design and usability test process of the system, the researcher logs were analyzed by content analysis. While analyzing the screen recording data obtained from the students, task completion times and task completion times in the cam studio program were examined. Think aloud data and interview data obtained from the students and interview data analysis from the experts were made with content analysis. The rubric data of the experts were calculated by scoring 1-5.

Usability of ArtiBos

In this study, the findings collected from experts and students are organized under two headings: design and usability. The data sources and data collection tools benefitted during the process are outlined in Figure 5.

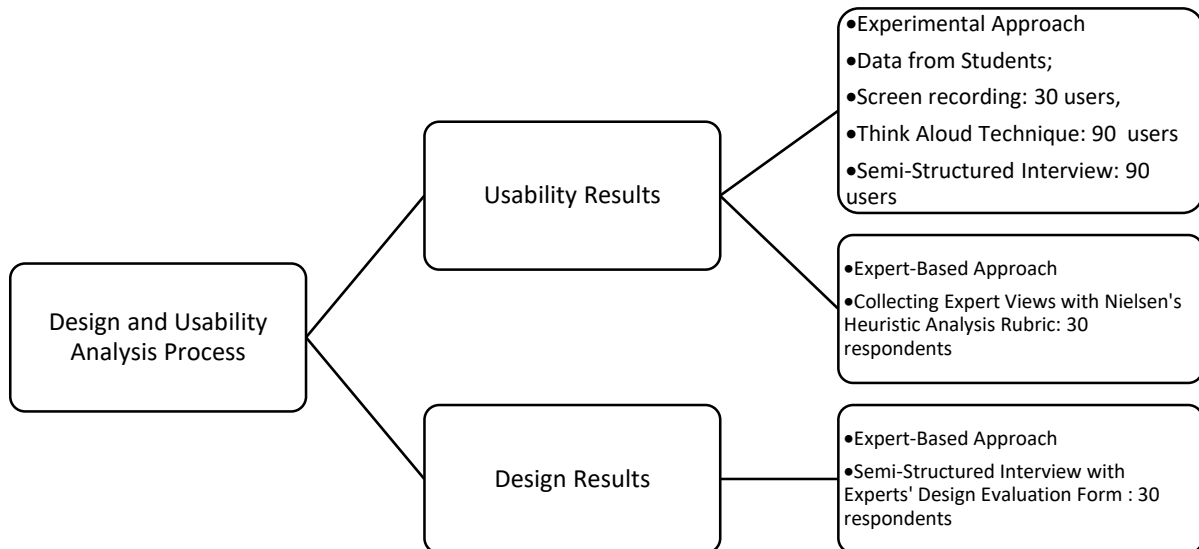


Figure 5. Process of design and usability analysis

As seen above, the results obtained from the implementation process are divided into two as design and usability. The tools indicated in Figure 5 were used while collecting data from the participating students and experts.

Usability Results

The usability of the system interface design was assessed by using experimental and expert-based approaches. For this reason, data were collected both during and at the end of the process.

Within the scope of experimental approach, usability tests and interviews were conducted with students who are the primary users of ArtiBos. As for expert-based approach, interviews were held with field experts to learn their evaluations about ArtiBos.

Expert-based approach was used during the practical implementation stage. Therefore, the experts kept questioning the usability of the system and required made necessary updates starting from the beginning. At the end of the process, both expert-based and experimental approaches were utilized. In this way, it became possible to evaluate the system from the perspective of both design experts and students as the main users.

Data from Students

In order to be able to evaluate the effectiveness and efficiency of the system, quantitative data on completion status and average completion time of the tasks were used, as set out in Table 4.

Table 4

Quantitative Data regarding Effectiveness and Efficiency of ArtiBos

Task	Task Completion Status (%)	Average Completion Duration (sec)
Task 1	76.6	51.3
Task 2	100	5.5
Task 3	90	16.3
Task 4	93.3	8.8
Task 5*	-	-
Task 6	93.3	320
Task 7	46.6	407
Task 8	100	6.5

* Evaluated with qualitative data.

Task 1: Update your e-mail account

It was found out that 23 out of 30 users could complete Task 1. The average completion time was recorded as 51.3 seconds. Below are quoted some comments made by users who failed to complete the task at all or do it on time. The notes were taken via think aloud and interviews:

“Where will I updated the e-mail, there is not an option.” (CAHS - S2)

“I cannot update my e-mail, teacher..” (CAHS - S5)

“The changes we write are not being saved” (YSSAHS - S8)

Task 2: Send a message

On task 2, all users were seen to quickly reach the message menu, which is conveniently accessible and visible by users at all levels. As an exception, the users were not able to send messages due to a technical problem that occurred on the day the data was collected.

Although users had no problems with the interface, they brought the following criticisms because the operation could not be performed:

“I don’t think it is writing a message.” (CAHS - S2)

“When I write a message, it gives an error, so I cannot send a message.” (CAHS - S11)

“It is not sending a message, teacher.” (FSMAHS – S4)

“Messaging gives an error.” (YSSAHS – S1)

Task 3: Access to the instruction content for mixture problems

It was seen that task 3 was completed by 27 students, while the rest of the 3 students did not start the task at all. The average duration of access to the instruction content by students was calculated as 16.3 seconds. Below are quoted some of the users’ comments noted during the implementation via think aloud technique:

“I enjoyed the lecturing. It is OK, I mean.” (CAHS – S5)

“Teacher, I can't go to the other page, the same stuff keeps blinking.” (CAHS – S8)

“How will I access to the instruction content? Via learn?” (CAHS – S20)

“How can one access to the instruction content?” (FSMAHS – S10)

“A bit complicated, I don’t understand where to click” (YSSAHS – S10)

Task 4: Watch the video tutorial on how to create the problem

It was seen that the video mentioned in the task and placed under the menu “Watch” on the system could be accessed by 28 of the students leaving only 2 unable to perform the task. The average duration of task completion was noted as 8.8 seconds. Below are noted some opinions regarding the task in question:

“In the place titled watch, something can be put, like learn to create a problem.”
(CAHS – s4)

“How did you play the video?” (CAHS – S10)

“Video texts are streaming fast, they should be slower.” (CAHS – S8)

“Video texts could be more active.” (CAHS – S5)

“It zooms in and out, it is clear I suppose.” (CAHS – S7)

“I think things in the video might be more active.” (CAHS – S4)

“I think the colours and so on are very lovely.” (FSMAHS – S5)

Task 5: Learn how to use the system with the wizard

As a result of completing this task, it is planned that students will have learnt how to use the system by means of the wizard. It was identified as a qualitative task in order to find out whether the learning outcome took place. The data concerning task 5 were obtained by analyzing the interviews and screen records.

The screen recordings show that when students first logged in the system, they tended to use it randomly and discover it on their own. However, those who noticed the wizard could reach the menus they want easily as a result of reading the wizard’s instructions. Also, they recommended using the wizard for those having difficulty in surfing the system. As an example, one student was heard to warn a peer as following:

“Look, it writes what it is next to the wizard.” (FSMAHS – S9)

“You can learn what you cannot do by looking at the wizard.” (YSSAHS – S9)

During the subsequent interviews, the students were asked for their opinions about the function and design of the wizard. Most of the students stated that the wizard was useful.

“Teacher, the wizard explains what they are for, so it helped me well.” (CAHS – S12)

Yet, there were some students who find the text field next to the wizard complex and the succession of the informations as confusing. Some of the comments and remedies mentioned in this regard are given below as an example:

“I think the part next to this wizard is unclear, teacher. For example, it would be better if it explained loudly when one gets into it. It's messing up now.” (CAHS – S1)

“If instructions for the menus are given within the menus, rather than through the robot, they will be more understandable. Instead of all of the information in a consecutive way, I think it would be better to give in the relevant menus.” (CAHS – S17)

It was seen that the students hold an overall positive view about the wizard design. In particular, it was pointed out that the colour and the design are nice besides having an appeal to the target group of users.

“Teacher, the colour of the wizard is beautiful in my opinion.” (FSMAHS - S1)

Reviews and recommendations of some students regarding the wizard design are quoted below:

“I think there might be something moving inside the character, teacher.” (FSMAHS - S3)

“The character could have been a bit flashier. For example, it may have a metallic look.” (FSMAHS – S7)

Task 6: “25 kg of salty water with 16% salt is mixed with 15 kg of salty water with a ratio of 4%. What is the resulting percentage of salt of the resulting mixture?” Create the problem.

All of the 30 students were observed to attempt the task, while 28 of them could successfully complete it. The average duration of building the problem was 320 seconds, which is longer than expected because the students used the system for the first time and they took their time to recall their previous information about mixture problems. Thus, it does not sound unfair to expect a decrease in the duration as students become more experienced in using the system

and the system interface is updated as required. Some of the students' criticisms about the problem creation page are quoted below:

"How will I add the substances?" (CAHS – S3)

"What will we select and create now." (YSSAHS – S8)

"I can't undo the thing I add." (FSMAHS – S4)

Task 7: Choose and solve a mixture problem

It was found out that while 14 students managed to complete the task, 16 others could not. The average task completion duration was saved as 407 seconds. Considering the data obtained from think aloud technique and interviews, these two findings can be explained with the assumption that the students had not experienced this particular kind of problem solving operation before, they had been expecting a multiple choice question as they were used to solve that kind of problems, and they were not able to fully understand the problem and the missing facts in the problem. The citations below might better highlight this case:

"The list of given and missing facts is too complicated." (YSSAHS-S7)

"I was confused by the existence of too many options." (YSSAHS-S15)

"At first I didn't understand anything, it was confusing, but when I read it, I understood and could solve it." (YSSAHS-S20)

"I was expecting a multiple choice test. I did not understand how to solve, it's very complicated for me." (YSSAHS-S21)

Task 8: Check out the help menu

It was found out that all of the participants could successfully complete this task. As a result of the usability assessments, a number of availability issues were detected in the system as listed in Table 5. Other issues are given in Table 6. In summary, 70 of the 90 evaluating students articulated several issues, while 20 people stated that the system was smooth.

Table 5

Usability Problems Encountered in Access to ARTIBOS

Usability Problems	n=70	%
Design		
Functions of menus are not explicitly stated	50	71.42
Spelling mistakes	7	10
Colour, visuality, font and size problems	29	41.42
Menu and button size and positions	21	30
Technical (hardware-related)		
Failure to send messages	48	68.57
Hardware-related technical	36	51.42

According to Table 5; among usability problems, the interface design problems are generally caused because some of the menus could not be understood by the primary users, the text in some menus are too small, and guidance is not enough. Though less important, there were found some problems with illustration and colouring such as inadequate images and pale colours. In relation with technical problems under usability, the first rank is seen to be occupied by the problem with sending messages. Although the positioning of the message button seems usable enough, messages could not be sent due to the technical problem on the day of the test. Technical problems such as the slow running of the web site and halting video were caused by the Internet infrastructure. Similarly, loudspeakers did not work because of the testing computers. As a result, it was planned to update the web site as needed to eliminate the problems with the web interface.

Table 6

Other Problems Encountered in Access to ARTIBOS

Problem	n=70	%
Content		
Difficulty in solving problems	14	20
Confusing weight units	32	45.7
Lecture module does not contain sufficient sample problem solutions	4	5.7

Other problems concerning the system were reported as difficulty of the questions, confusing the weight units, and lack of sufficient sample questions. In order to solve these problems, the mathematics experts changed the content and necessary updates were made in the system.

Expert Views

The evaluation findings of the rubric prepared according to the design experts and Nielsen (1994) heuristics are shown in Table 7.

Table 7

Expert Evaluation

Heuristics	Average
Users are able to receive instant notification about where they are and what they do as long as they are online.	2.43
Information in the system is expressed with familiar and common concepts, text, and pictures.	3.90
The system does not restrict user freedom, so users are at liberty to log out or restart the system whenever they wish.	3.30
The system is standard and consistent in itself. For example, a specific design and font is used throughout the system.	3.80
Possible errors in the system are estimated and taken precaution. For example, when an object is deleted, it is done after the confirmation as “Are you sure that you want to delete this item?”	3.23
When the user is using the system or switching from one place to another, s/he does not have to remember the previous part.	2.83
The system can be easily used by users of all levels.	2.86
Error messages in the case of system errors while using of the system are expressed with a simple language for everybody.	3.23
The system has a help menu, which is focused on solving problems that might come up while using the system.	3.70
The system is aesthetic and simple.	3.73

The rubric based on Nielsen (1994) heuristics consists of five options for each item as “Totally disagree“, ”Disagree“, ”Not sure“, “Agree”, and “Completely agree”. The items are scored between 1 and 5. When the average score values of the answers given by the experts are considered, the lowest score is seen to refer to two heuristics, which are “Users can get instant feedback in the system” and “The system can be used comfortably by users at all levels”. Conversely, the heuristics referring to the system interface were rated highly:

“Functional help menu of the system”, “Simple and aesthetics interface design”, and “Expressing the content with familiar pictures and concepts”.

Design Results

After the experts’ evaluation based on Nielsen’s (1994) heuristic rubric, the design evaluation form of open-ended questions was responded by the same experts for in-depth discussion of the results. The frequency of answers for each item is given in Table 8.

Table 8

Frequency of Problems Referred in Design Evaluation

Problem	n=30
Colour, size and positioning problems concerning text and images	10
Lack of feedback	3
Screen usage, alignment, and symmetry problems	4
Object problems used during problem-building	9
Functionality problems regarding processes used during problem solving	8
Deficiencies on the personal information page	3

The table above implies indicates that the criticism about the design of the web site are attributed to text and images, feedbacks, screen usage, alignment and symmetry, objects used in problem-building, operations used in problem solving, and menus on the personal information page. Some of the opinions expressed by experts about design problems are cited as follows:

“The feedback given to the questions is not enough. Sufficient feedback should be given for correct and incorrect answers.” (U5)

“Telephone number can be added to the profile menu. It can work for collective messaging.” (U21)

“The font size should be increased or an option should be added for customizing.” (U18)

“When solving a problem, feedback must be given after entering the given-missing facts.” (U2)

“What is written next to the main container when writing a problem is not clear. It can be within the box.” (U7)

Necessary updates were made on the system in the light of the criticism brought by the participants.

Discussion and Conclusion

In this study, the usability evaluation process is designed for AITs. Firstly, the usability evaluation processes of ITSs and AITs were examined. Then, it was explained that what is done in the usability evaluation process of Arti Bos. Finally, an improved usability evaluation process was applied. The data obtained in the study were discussed for the usability evaluation process.

In seeking answer for research question "How can the process of design and usability tests be planned for AITs?", exhaustive research was carried out about the features, pros and cons of the usability test approaches including the review of the literature on usability studies along with their results. The review of the existing literature shows that usability of digital educational environments, intelligent tutoring systems, and adaptive teaching systems are mostly with experimental method (Beymer, Orton, & Russell, 2007; Bayram & Yeni, 2011; Weber & Brusilovsky, 2001; Dimitrova, 2003; Lin, Wu, & Hsueh, 2014; Dexheimer, et al., 2017; Pala, Arslan, & Özdiç, 2017; Ramirez-Norigea, Juarez-Ramirez, & Martinez-Ramirez, 2017; Erdoğan & Şahin, 2018). It is seen that other usability studies targeting the same type of systems are conducted with expert-based approach (Sezer, 2011; Trintafillou, Pomportsis, & Georgiadou, 2003). However, there seems a scarcity of studies combining both approaches (Roscoe, et al., 2004; Verkuyl, et al., 2016).

In this study, the usability of Artibos was carried out with both user-based and expert-based approaches. Consequently, highly detailed and inclusive findings were yielded. Furthermore, the results are considered even more realistic because the user-based approach was used in the normal classroom with the main users of the system in the natural classroom environment. Although the literature suggests that usability test with expert-based approach could unveil most problems concerning the system, an application involving the real users of the system conducted with user-based approach seems highly promising for appraising effects of the problems (Hollingsed, 2007). This combines the advantages of both approaches

since expert-based approach is capable of predetermining problems and the other approach identifies effects of the problems found previously (Verkuyl, et al., 2016).

As a result, it can be said that design of such systems by both design experts and the evaluation of the system through eyes of end users would increase the usability of the system. Another strength of our usability test is the rationale of the rubric given as a part of expert-based approach, that is Nielsen's (1994) heuristics as one of the most established grounds in this area. Moreover, the advantages of expert-based approach as shorter time of application and practicality of reporting were enjoyed in this study (Allen, 2006). This implies that researchers had better decide on the research approach through a comprehensive fact-finding trial of not only study participants but also the advantages and disadvantages of the approach such as time and cost.

Apart from the testing approaches, the present study on the usability of Artibos also resembles a potent example as it combines in-process and end-of-process test types. In-process testing proved helpful in spotting and correcting design problems from the very beginning of the process. Thanks to this aspect, problems that are normally difficult to correct later could be overcome more easily and systematically. As for the end-of-process test, problems that were ignored in the process were exposed and removed. It can be inferred that mutual use of the testing models offers a significant advantage. Çağıltay (2011, p. 98) also stated that both types should be used together for more effective results.

Departing from the logs carefully kept during the design and usability assessment of the system, we would like to add the following recommendations for the design evaluation of a similar system in the future:

- Before applying the test, a letter of consent should be obtained from all participant students' parents and also the school of data collection in case of any complaints. The ethics committee report should also be enclosed.
- It should be made sure that computers to be used in testing and facilities such as the Internet connection, mouse, keyboard, camera, etc., are fully in operating state.

- Participants should be selected on a complete volunteer basis, reminding their liberty to leave the test at any time they wish, and substitute users should be identified just in case of drop-outs.
- Due diligence should be shown to time planning as it is a complicated stage to select users, obtain the necessary permissions, and prepare the test environment.
- It would be in the interest of researchers to make the testing environment and equipment ready in a timely manner in case of down-time in data collection.

References

- Acartürk, C., & Çağıltay, K. (2006). İnsan bilgisayar etkileşimi ve ODTÜ'de yürütülen çalışmalar. *Akademik Bilişim*, 335-340.
- Al-Bastami, B. G., & Naser, S. S. A. (2017). Design and development of an intelligent tutoring system for c# language. *European Academic Research*, 4(10), 8795-8809
- Allen, M., Currie, L.M., Bakken, S., Patel, V.L., & Cimino, J.J. (2006). Heuristic evaluation of paper-based web pages: a simplified inspection usability methodology. *Journal of biomedical informatics*, 39(4), 412-423.
- Arevalillo-Herráez, M., Marco-Giménez, L., Arnau, D., & Calero, J.A.G. (2017). Adding sensor-free intention-based affective support to an Intelligent Tutoring System. *Knowledge-Based Systems*, 132, 85-93.
- Arsoy, S., Kalıpsız, O., & Öztürk, S. (2013). *E-devlet web siteleri için rehber tabanlı kullanılabilirlik değerlendirmesi*. Yedinci Ulusal Yazılım Mühendisliği Sempozyumu, 1-9.
- Bailey, G.S. (1993, May). *Iterative methodology and designer training in human-computer interface design*. In Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems (pp. 198-205). ACM.
- Baylari, A., & Montazer, G.A. (2009). Design a personalized e-learning system based on item response theory and artificial neural network approach. *Expert Systems with Applications*, 36(4), 8013-8021.
- Bayram, S., & Yeni, S. (2011). Web tabanlı eğitsel çoklu ortamların göz izleme tekniği ile kullanılabilirlik açısından değerlendirilmesi. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, 12(2), 221-234.
- Benyon, D. (1993). Adaptive systems: A solution to usability problems. *User modeling and User adapted Interaction*, 3(1), 65-87.
- Bernacki, M.L., Aleven, V. & Nokes-Malach, T.J. (2014). Stability and change in adolescents' task specific achievement goals and implications for learning mathematics with intelligent tutors. *Computers in Human Behaviour*, 37, 73-80.

- Bernacki, M.L., Aleven, V. & Nokes-Malach, T.J. (2014). Stability and change in adolescents' task specific achievement goals and implications for learning mathematics with intelligent tutors. *Computers in Human Behaviour*, 37, 73-80.
- Beymer, D., Orton, P.Z., & Russell, D.M. (2007). *An eye tracking study of how pictures influence online reading*. IFIP Conference On Human Computer Interaction, 456-460.
- Buscher, G., Dumais, S.T., & Cutrell, E. (2010). *The good, the bad, and the random: an eye tracking study of ad quality in web search*. 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 42-49.
- Can, G.F., Atalay, K.D., & Eraslan, E. (2017). Tabletlerin kullanılabilirlik ölçütlerine göre çok kriterli karar verme yaklaşımıyla değerlendirilmesi. *Mühendislik Bilimleri ve Tasarım Dergisi*, 5(SI), 81-88.
- Card, S. K. (2017). *The psychology of human-computer interaction*. CRC Press.
- Carroll, J. M. (2003). Introduction: towards a multidisciplinary science of human- computer interaction. *HCI Models, Theories and Frameworks*, 1-9.
- 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.
- Cevher, E. (2015). Kamu üniversiteleri web sayfalarının kullanılabilirliğinin içerik analizi ile incelenmesi. *University of Gaziantep Journal of Social Sciences*, 14(2), 387-402.
- Chen, C.M., & Duh, L.J. (2008). Personalized web-based tutoring system based on fuzzy item response theory. *Expert Systems with Applications*, 34(4), 2298-2315.
- Chen, C.M., Lee, H.M., & Chen, Y.H. (2005). Personalized e-learning system using item response theory. *Computers & Education*, 44(3), 237-255.
- Chughtai, R., Zhang, S., & Craig, S.D. (2015). Usability evaluation of intelligent tutoring system: its from a usability perspective. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59(1), 367-371.
- Cockrell, B. J., & Jayne, E. A. (2002). How do I find an article? Insights from a web usability study. *The Journal of Academic Librarianship*, 28(3), 122-132.

- Cutrell, E., & Guan, Z. (2007). What are you looking for?: an eye-tracking study of information usage in web search. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 407-416.
- Çağiltay, K. (2011). *İnsan bilgisayar etkileşimi ve kullanılabilirlik mühendisliği: teoriden pratiğe*, ODTÜ Yayıncılık, Ankara.
- Çebi, S., Durucu, M., & Kayhan, B. M. (2013) *gençlerin internet kullanım alışkanlıkları ve kullanılabilirlik üzerine bir çalışma: kütüphane web sayfalarının kullanılabilirliği*. 19. Ulusal Ergonomi Kongresi, 525-534.
- Çetin, İ., & Şendurur, E. (2016). Çevrimiçi akademik kaynakların kullanılabilirlik değerlendirmesi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 1(40), 273-299.
- Dexheimer, J.W., Kurowski, B. G., Anders, S.H., McClanahan, N., Wade, S.L., & Babcock, L. (2017). Usability evaluation of the smart application for youth with mTBI. *International Journal of Medical Informations*, 97, 163-170.
- Dimitrova, V. (2003). STyLE-OLM: Interactive open learner modelling. *International Journal of Artificial Intelligence in Education*, 13(1), 35-78.
- Doğan, N., & Kubat, B. (2008). Zeki öğretim sistemleri için yeni bir bileşen: düzenleyici modül. *International Journal of Informatics Technologies*, 1(2). 5-9.
- Dönmez, O., Yaman, F., Şahin, Y.L., & Yurdakul, I.K. (2016). İşitme engelliler için mobil uygulama geliştirme süreci: Çarkifelek örneği. *Eğitim Teknolojisi Kuram ve Uygulama*, 6(1). 20-41.
- Dumas, J.S., Dumas, J.S., & Redish, J. (1999). *A practical guide to usability testing*. Intellect books.
- Durmus, S., & Çağiltay, K. (2012). Kamu Kurumu Web Siteleri ve Kullanılabilirlik. E-devlet. M.Z. Sobacı, M. Yıldız (Ed.), *Kamu Yönetimi ve Teknoloji İlişkisinde Güncel Gelişmeler* içinde (ss. 293-322). Nobel Yayınevi.
- Erdoğan, F., & Şahin, S. (2018). Her yerde öğrenme sisteminin kullanılabilirliğine ilişkin öğrenci görüşleri. *Karaelmas Eğitim Bilimleri Dergisi*, 6(1).
- Gagneux, A., Eglin, V., & Emptoz, H. (2001). Quality approach of web documents by an evaluation of structure relevance. *Proceedings of WDA*, 11-14.

- Hertzum, M., Hansen, K. D., & Andersen, H. H. (2009). Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, 28(2), 165-181.
- Hollingsed, T., & Novick, D.G. (2007, October). *Usability inspection methods after 15 years of research and practice*. In Proceedings of the 25th annual ACM international conference on Design of communication (pp. 249-255). ACM.
- Hooshyar, D., Ahmad, R.B., Yousefi, M., Fathi, M., Horng, S.J., & Lim, H. (2016). Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. *Computers & Education*, 94, 18-36.
- Iqbal, M., & Ullah, A. (2016). Usability evaluation of hec national digital library website: a qualitative approach. *Library Hi Tech News*, 33(3). 8-10.
- ISO 9241-11 (1998). Ergonomic Requirements for Office Work with Visual Display Terminals (Vdts) Part 11: Guidance on Usability. *International Organization for Standardization, Geneva*.
- ISO/IEC 9126-1 (2001). Software Engineering Product Quality -- Part 1: Quality Model. *International Organization for Standardization, Geneva*.
- Jaspers, M.W.M. (2009). A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. *International Journal of Medical Informatics*, 78, 340-353.
- Jovanović, J., Gašević, D., & Devedžić, V. (2006). Dynamic assembly of personalized learning content on the semantic web. In *European Semantic Web Conference* (pp. 545-559). Springer, Berlin, Heidelberg. Kamis (2018). https://kamis.gov.tr/wpcontent/uploads/rehber/Rehber_Bolum6_KullanilabilirlikTestleriveUyulamalari.pdf adresinden 17 Temmuz 2018 tarihinde alınmıştır.
- Khalfallah, J., & Slama, J.B.H. (2015). Facial expression recognition for intelligent tutoring systems in remote laboratories platform. *Procedia Computer Science*, 73, 274-281.
- Koedinger, K.R., & Corbett, A. (2006). *Cognitive Tutors: Technology Bringing Learning Sciences to the Classroom*.
- Kularbphetpong, K., Kedsiribut, P., & Roonrakwit, P. (2015). Developing an adaptive web-based intelligent tutoring system using mastery learning technique. *Procedia-Social and Behavioral Sciences*, 191, 686-691.NA.

- Lazar, J., Feng, J., & Hochheiser, H. 2010. *Research Methods in Human Computer Interaction*. Morgan Kaufmann.
- 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.
- Lemke, C. (2013). Intelligent adaptive learning: an essential element of 21st century teaching and learning. Bellevue, wa: dreambox learning. <http://www.dreambox.com/white-papers/intelligent-adaptive-learning-an-essential-element-of-21st-century-teaching-and-learning>. adresinden 20.12.2018 tarihinde alınmıştır.
- Lin, H.C.K., Wu, C.H., & Hsueh, Y.P. (2014). The influence of using affective tutoring system in accounting remedial instruction on learning performance and usability. *Computers in Human Behaviour*, 41, 514-522.
- Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and intelligent systems for collaborative learning support: A review of the field. *IEEE transactions on Learning Technologies*, 4(1), 5-20.
- Myers, B.A., & Rosson, M.B. (1992). Survey on user interface programming. *SIGCHI Conference on Human Factors in Computing Systems*, 195-202.
- Nielsen, J. (2012). Usability 101: Introduction to usability. Retrieved 14, August, 2018, from <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>.
- Norman, D.A. (1988). *Designing every day things*. Currency-Doubleday, New York.
- Olson, J.R., & Olson, G.M. (1995). The growth of cognitive modeling in human-computer interaction since goms. In *Readings in Human-Computer Interaction* (pp. 603-625).
- Oyibo, K., Ali, Y.S., & Vassileva, J. (2016). An empirical analysis of the perception of mobile website interfaces and the influence of culture. *International Workshop on Personalization*, 44-56.
- Özdemir, S., Atasoy, B., & Somyürek, S. (2007). Bilimsel dergilerin iş süreçleri yönetimini gerçekleştiren bir yazılımın kullanılabilirlik araştırması: türkiye'deki ilk örneğin incelenmesi. *Gazi Universtiy Journal of Gazi Educational Faculty* 27(2), 57-80.
- Pala, F.K., Arslan, H., & Özdiñç, F. (2017). Eğitim bilişim ağı web sitesinin otantik görevler ve göz izleme ile kullanılabilirliğinin incelenmesi. *Ihlara Eğitim Araştırmaları Dergisi*, 2(1).

- Papanikolaou, K.A., & Grigoriadou, M. (2003). *An instructional framework supporting personalized learning on the Web*. In *Advanced learning technologies, 2003. Proceedings. The 3rd IEEE international conference on* (pp. 120-124). IEEE.
- Phobun, P., & Vicheanpanya, J. (2010). Adaptive intelligent tutoring systems for e-learning systems. *Procedia-Social and Behavioral Sciences*, 2(2), 4064-4069.
- Radenkovic, B.L., Despotovic-Zrakic, M.S., Barac, D.M., Bogdanovic, Z.M., & Milic, A.R. (2011). *Web portal for adaptive e-learning*. In *Telecommunication in Modern Satellite Cable and Broadcasting Services, 10th International Conference*, 365-368.
- Ramirez-Norigea, A., Juarez-Ramirez, R., & Martinez-Ramirez, Y. (2017). Evaluation module based on bayesian networks to intelligent tutoring systems. *Journal of Information Management*, 37(1), 1488-1498.
- Reilly, E. D. (2003). *Milestones in computer science and information technology*. Greenwood Publishing Group: Connecticut, USA.
- Reiners, T., & Dreher, H. (2009). Culturally-based adaptive learning and concept analytics to guide educational website content integration. *Journal of Information Technology Education*, 8(1), 125-139.
- Roscoe, R.D., Allen, L.K., Weston, J. L., Crossley, S.A. & McNamara, D.S. (2014). The writing pal intelligent tutoring system: usability testing and development. *Computers and Composition*, 34, 39-59
- Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests*. John Wiley & Sons.
- Sanchez, R.P., Bartel, C.M., Brown, E., & DeRosier. M. (2014). The acceptability and efficacy of an intelligent tutoring system. *Computers&Education*, 78, 321-332.
- Sezer, İ. (2011). Hipermedya sistemlerinde uyarlanabilir ve uyarlanırlar metodları karşılaştırma ve yabancı dil öğretiminde örnek bir araç geliştirme. *Yüksek Lisans Tezi, Gazi Üniversitesi, 100s (yayınlanmamış)*. (Tez No: 285490).
- Shneiderman, B. (2010). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Education India.
- Siemer, J., & Angelides, M.C. (1998). A comprehensive method for the evaluation of complete intelligent tutoring systems. *Decision Support Systems*, 22(1), 85-102.

- Triantafillou, E., Pomportsis, A., & Georgiadou, E. (2003). AES-CS: adaptive educational system based on cognitive styles. *Computers & Education*, 41(1), 87-103.
- Tuna, G., & Öztürk, A. (2015). *Zeki ve uyarlanabilir e-öğrenme ortamları*. International Distance Education Conference 2015 – September 2-4, St. Petersburg, Russia
- Verkuyl, M., Atack, L., Mastrilli, P., & Romaniuk, D. (2016). Virtual gaming to develop student's pediatric nursing skills: a usability test. *Nurse Education Today*, 46, 81-85.
- Victorio-Meza, H., Mejia-Lavalle, M., & Ortiz, G.R. (2014). *Advances on knowledge representation of intelligent tutoring systems*. In Mechatronics, Electronics and Automotive Engineering International Conference, 212-216.
- Vos, H. J. (1995). Applications of bayesian decision theory to intelligent tutoring systems. *Computers in Human Behaviour*, 11(1), 149-162.
- Weber, G., & Brusilovsky, P. (2001). ELM-ART: An adaptive versatile system for Web-based instruction. *International Journal of Artificial Intelligence in Education (IJAIED)*, 12, 351-384.
- Yavuz, İ., Çınar, N.Ö., & Çağiltay, K. (2016). Kamu internet sitelerinde yer alan arama alanlarının kullanılabilirliği ve buna yönelik kullanıcı davranışlarının belirlenmesi. *International Journal of Informatics Technologies*, 9(1), 41.
- Zviran, M., Glezer, C., & Avni, I. (2006). User satisfaction from commercial web sites: the effect of design and use. *Information & Management*, 43(2), 157-178.