

A Hybrid Multi-Dimensional Usability Evaluation Approach Proposal: A Case Study Related to Airline Firms' Websites

Gülin Feryal Can (✉ gfcan@baskent.edu.tr)

Baskent University

Elif KILIÇ DELİCE

Ataturk University: Ataturk Universitesi

Research Article

Keywords: Usability, PSI, ENTROPY, CRITIC, CODAS

Posted Date: April 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-420254/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A HYBRID MULTI-DIMENSIONAL USABILITY EVALUATION APPROACH PROPOSAL: A CASE STUDY RELATED TO AIRLINE FIRMS' WEBSITES

Gülin Feryal CAN*

Başkent University, Engineering Faculty, Industrial Engineering Department, Ankara,
Turkey, gfcan@baskent.edu.tr

Elif KILIÇ DELİCE

Atatürk University, Engineering Faculty, Industrial Engineering Department, Erzurum,
Turkey, elif.kdelice@atauni.edu.tr

Abstract

This study aims to advance a new multi-dimensional usability evaluation approach (MDUEA) for websites including four phases. In the first phase, components of usability evaluation as tasks, participants, usability criteria are determined. In the second phase, the Morae V3 package program used for the tests recording the user actions and the system usability scale (SUS) is performed by Morae at the end of the each task. In the third phase, Preference Selection Index (PSI), ENTROPY and The Criteria Importance through Inter criteria Correlation (CRITIC) are performed to obtain objective usability criteria weights considering a novel three-dimension initial decision matrices than, the weights obtained from these three techniques are aggregated via geometric mean operator. In the same phase, Rank Sum (RS), Rank Reciprocal (RR) and Rank Exponent (RE) methods are utilized to compute subjective criteria weights than, weights obtained from these three approaches are aggregated via using geometric mean operator. In the fourth phase, Combinative Distance-based Assessment (CODAS) is used to rank websites considering a new proposed threshold function value computation for objective criteria group and subjective criteria group separately.

Keywords: Usability, PSI, ENTROPY, CRITIC, CODAS

1. INTRODUCTION

The usability of websites basically depends on their ease of use, achievement of the purpose in a short time and ensuring user satisfaction. Websites that fail to meet these conditions are left by the users and another website, that is, another product, company, service, etc. is preferred. Especially in commercial websites, a website not preferred by users means loss of customer.

The low level of usability in non-commercial websites, such as course management systems, library websites, train and bus websites etc. prevents the intensive and active usage of such websites, as it affects the complexity and activity time that users spent. This situation brings with activities such as resource sharing and communication, which are carried out for longer time periods outside the web environment and with loss of information.

Different users can access a website for different purposes. Level of users' knowledge related using a website may vary due to criteria such as their ages, experiences, hardware used and web browsers. Accordingly, a website should be able to reach the user purpose without any problem to cover different user features. However, in this condition, the website will be able to meet the needs of users for being successful according to view points of users. In order to ensure the usability of the websites, clear and understandable designs to make the user think very little in order to reach the purpose. It is important for usability that a website is accessible, aesthetic, consistent, open, simple, and navigable and to support inexperienced users.

Different techniques can be used to evaluate the usability of a website. However in general, these techniques can be divided into four categories. These categories are empirical methods, expert reviews, user tests and biometrics analysis. Empirical methods aim to collect and quantify the opinions of people using the software system (Nielsen, 1993). These techniques include survey and thinking aloud methods. Survey techniques are one of the most performed methods in usability analysis. There are many specialized surveys to examine different types of systems. Among these, Software Usability Measurement Inventory (SUMI) and Website Analysis Measurement Inventory (WAMMI) are very effective in measuring the usability of software and web sites (Nielsen, 1993; Kirakowski et al., 1998). The technique of thinking loud is based on the expressing user's views about the system, his difficulties and the points he likes while using the software or website. Today, this technique is used in the evaluation of video games where the survey technique is not sufficient (Nielsen, 1993; Kirakowski et al., 1998). Another group of techniques applied in usability analysis is expert evaluations. These are based on the fact that one or more experts with experience in the field of usability examine the system and identify possible user errors and possible misunderstandings (Nielsen, 1993). Within the scope of expert evaluations, heuristic evaluation and conceptual trend methods are applied. Heuristic evaluation method introduced by Nielsen (1993) summarizes usability parameters in 10 rules. Among of them, 7 rules are related to the design. These are "flexibility and efficiency", "user control and freedom", "continuity and standards", "aesthetic and minimalist design", "understanding instead of remembering", "visibility of the system state, matching system and

real life". These rules represent effectiveness, efficiency, learnability and satisfaction criteria. The 3 rules declared as "error prevention", "help users understand, identify and avoid mistakes", "help and documentation" are related to the help criteria. The expert, who conducts the usability evaluation through a checklist, evaluates the system within the framework of these rules (Nielsen, 1993). Conceptual trend method is based on the prediction of an expert or expert group that is not involved in the design phase, determining the possible tasks that the user may want to perform on the system, and foresee the problems and mistakes that the user may encounter during the task performance. In order to perform conceptual trend methodology, it is necessary to first define the users and scenarios, and then to determine the possible actions of the users in the scenario stages. Then, the expert tries to identify the possible mistakes and errors of the users by proceeding through the stages of action (Lewis & Wharton, 1997). The third method group used in usability analysis is user tests. In user tests, a user group targeted by the software system is asked to perform the tasks specified in previously created scenarios using the software. According to Nielsen (1993), 5 participants may be sufficient to test such systems and these participants can detect existing errors by 75%. When the number of subjects increases, the number of errors detected does not increase significantly. Within the scope of the method, the paths followed while using the system for each user are observed and measured together with the wrong and correct processing times. For this purpose, it is possible to benefit from pre-prepared interface and process flow diagrams, as well as video recordings and special software (Nielsen, 1993). The fourth method group is biometrics analysis. In biometrics analysis, based on the user's physical reactions; usability is tried to be evaluated. This method is used together with user tests (Nielsen, 1993). Within the scope of biometrics analysis, eye-tracking and EEG methods are used. In the eye tracking method, the points that the user focuses on the interface are determined; the movement of the eye between these points and the focusing time at these points are measured. Special hardware and software are required for the application of eye tracking, and the interpretation of the obtained data requires extensive expertise (Nielsen, 1993). In the EEG method; EEG device that measures the electrical activity of the brain is used. However, it is not a widely used method because of the cost arising from the need for special hardware and software, the difficulty of analyzing the data obtained, and the effect of physical responses other than cerebral activity (Faulkner, 2003). Among these methods under four categorizes, user tests are expressed as the most common used usability assessment methods performed with real users (Nielsen, 1993).

When examine the methods used for usability evaluation, it is seen that each of them has a different perspective. While some methods work with subjective data, some methods work with numerical data. Additionally, each method has some disadvantages and advantages in terms of application cost, application time, required education level, level of success in determining problems, etc. comparing each other. In some studies, different usability evaluation methods have been performed and the results obtained from these different methods have been tried to associate. In addition, usability analysis was carried out by considering certain criteria in the studies. However different criteria as dimensions can affect system usability. In technical term, the influence of experts is at the highest level in some methods. Making usability evaluation should include both experts' opinions and software aided measurements in real time. More than one system has not been compared with different usability evaluation approaches at the same time in the literature and there is no study performed this integrated comparison based on Multi-criteria Decision Making (MCDM) structure. However, selecting the most usable system is one of the MCDM problem which has alternatives as compared systems, criteria as usability dimensions and decision makers as end users. When all these feedbacks are evaluated, it is seen that there is a requirement for a multidimensional approach that can carry out usability evaluation both from the perspective of the users and measured real data. For this reason, it is a hot topic performing MCDM approaches for usability evaluation. The usability evaluation techniques mentioned above are not sufficient for this type analysis. These techniques can be performed for each system separately and the results cannot be combined. For these reasons, it was aimed to propose a new multi-dimensional usability evaluation approach (MDUEA) for webpages in this study.

The proposed MDUEA includes four phases as *determining components of usability assessment, usability testing with MORAE V3, computing usability criteria weights by implementing different weighting approaches for objective and subjective usability criteria and aggregating weight results of these approaches, ranking webpages by performing Combinative Distance-based Assessment (CODAS) method*. In the first phase, components of usability evaluation as tasks, participants, usability criteria are determined. In the second phase, the Morae V3 package program used for the tests record the user actions. It can automatically provide standard measurement data such as mouse clicks, mouse movement distance, task time. The System Usability Scale (SUS) is performed by Morae software at the end of the each task. In the third phase, Preference Selection Index (PSI), ENTROPY and The Criteria Importance through Intercriteria Correlation (CRITIC) methods are performed to obtain usability criteria

weights than the weights obtained from these three techniques are aggregated via geometric mean operator. In the fourth phase, CODAS is used to rank websites considering three-dimension initial decision matrices for two criteria group as objective criteria group (OCG) and subjective criteria group (SCG).

The proposed MDUEA considers two different criteria group as OCG and SCG. OCG includes task time, time between two data entries, number of mouse clicks, mouse movement distance, and success percentage. In this way, efficiency, effectiveness and satisfaction dimensions were examined for the usability of 5 airline firms' web pages (AFWSs). Efficiency dimension is determined via measuring task time. While examining the effectiveness dimension, time between two data entries, number of mouse clicks and mouse movement distance are computed. Additionally, System Usability Scale (SUS) was applied to evaluate satisfaction dimension by measuring success percentage. SCG contains level of willingness to use, level of complexity, level of ease usage, required technical support level, feeling safe while using, level of required prior knowledge, level of integration goodness, level of inconsistency, learning speed, and cumbersome level.

Participants evaluated 5 AFWSs according to OCG and SCG via implementing different tasks. The purpose of assigning tasks in the suggested approach is to obtain accurate feedbacks related to the usability of AFWSs. This feedback is obtained by filling out the SUS for the AFWSs after the relevant tasks are completed by the participants. Each AFWS's performance for SCG is measured via ranking performed by participants for criteria in SCG from the most important to the least important and each AFWS's performance for OCG is computed with Morae V3 package program. Today, there are many programs that perform usability tests. The most common used analysis for usability tests was determined as the Morae software. The Morae software is also recommended by Jacob Nielsen, who is considered a usability expert. Nielsen was stated that "User testing has become much more efficient since Morae has been on the market" to emphasize the importance of performing usability tests with the Morae program (TechSmith, 2007).

In the proposed approach, the results of usability testing with Morae V3 package program and the results of SUS were combined by using MCDM structure. This combination has been carried out for the first time in the literature. In this context, Preference selection index (PSI), Entropy and CRITIC methods were used to obtain weight of OCG related to the usability of web pages. Then, by performing geometric average operator, the weights obtained from PSI, Entropy and CRITIC were aggregated. In this way, final weights of usability criteria in OCG

could be computed. This calculation was carried out objectively, without decision-makers' evaluations. In PSI developed by Maniya and Bhatt (2010), it is not necessary to assign a relative importance for criteria. PSI is useful when there is a conflict in deciding the relative importance among criteria (Attri and Grover, 2015). However, it can be used regardless of the number of criteria. The entropy is an effective method that can be used in different decision-making problems and also to measure the amount of useful information in a data set. It is highly reliable for information measurement and provides high accuracy in criteria weight determination problems. It is based on the view that the number or quality of information obtained from the decision-making environment affects the accuracy and reliability of the decision-making problem solution (Wu et al., 2011). Another method, Criteria Importance Through Intercriteria Correlation (CRITIC) can determine criteria weights objectively and it makes this determination easy for the decision maker to express the relative importance of the criteria (Diakoulaki et al., 1995). In CRITIC, criteria weights are obtained by focusing on correlation level and variability between the performance values of the alternatives. In the MDUEA, the weights of SCG were computed via using ranking based weighting methods as Rank Sum (RS), Rank Reciprocal (RR) and Rank Exponent (RE) developed by Malczewski (1999). In the context of MDUEA, criteria in SCG were evaluated by asking participants to place these criteria in an order of importance from 1 to 10 according to their opinions. Then, these ranks were used for RS, RR and RE procedures to obtain criteria weights for SCG. Finally, obtained criteria weights were aggregated by geometric average operator.

The second phase of the proposed MDUEA is ranking AFWSs in term of usability by performing CODAS advanced by Ghorabae et al. (2016). CODAS determines the selection index of an alternative via considering the Euclidean distance as the primary and the Taxicab distance as the secondary measure from the negative ideal point. The alternative which is further away from the negative ideal point is the more preferred one in CODAS. In the proposed MDUEA, it is aimed to determine which web site is the farthest from the web site with the least usability. In CODAS, if there are two alternatives which are incomparable for the Euclidean distance, the Taxicab distance is utilized as secondary measure. Since, it uses two different selection strategies, the rankings determined with CODAS is more logical than other methods such as WASPAS, COPRAS, EDAS, TOPSIS, VIKOR). Additionally, in this study traditional CODAS was advanced in term of its threshold function. Differences between Euclidean distances of two alternatives were compared according to standard deviation of Euclidean distances of all alternatives. In traditional CODAS, this comparison has been made according

to the threshold value assigned from decision maker between 0.01 and 0.05. The proposed standard deviation strategy can overcome this subjectivity.

The initial decision matrix for OCG was formed using objective data obtained from Morae V3 package program and the initial decision matrix for SCG was structured by implementing SUS for CODAS. Participants were applied SUS including 10 phrases related to each criterion in SCG for each AFWS. Participants gave point between 1 (strongly disagree) and 5 (strongly agree) for each phrase. These two initial decision matrices are three-dimension matrices including performance values of AFWSs for each criterion for each task. In this study, the initial decision matrix as three-dimension matrix has been advanced at first in the literature for task-based usability evaluation. This matrix includes tasks, websites and usability criteria at the same time. It has performance values of websites for usability criteria for each task.

Organization of remain sections is as follows. Second section includes literature review related to different websites usability and usability of AFWSs. Third section consists of the proposed MDUEA and its application for AFWSs, fourth section covers results and conclusions, and fifth section includes discussions and future research recommendations.

2. LITERATURE REVIEW

In this section the literature review study is given by separating usability evaluation in two groups as literature review including MCDM based usability evaluation and literature review related to AFWSs' usability evaluations.

Studies on website usability evaluation by using MCDM techniques have been increasing in recent years. For example, Tsai, Chou and Lai (2010) proposed an effective model for evaluating national park websites. In this model, the Decision-Making Trial and Evaluation Laboratory (DEMATEL) was applied to cope with the interdependencies between evaluation criteria while the Analytic Network Process (ANP) was used to determine criteria weights. In the same study, VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) was performed to rank Taiwan national park web page. Storto (2013) presented an integrative framework based on Data Envelopment Analysis (DEA) to evaluate e-commerce website efficiency respect to the user viewpoint. The framework was applied to compare 52 e-commerce websites, 9 variables were measured to classify websites in a set of 3 website macro-dimensions (user experience, site navigability and structure). A stepwise regression was performed to assess the impact of cognitive costs and benefits that often affect website productivity. Cebi (2013) supposed an integrated model using Delphi and DEMATEL techniques to present the

importance of websites design parameters based on different website types including interactions between these parameters by considering 3 shopping mall websites. Delphi method was used to determine website design parameters for website type and DEMATEL method was performed to determine importance of parameters. Karkin and Janssen (2014) evaluated 16 websites of Turkish metropolitan municipalities from the conventional and public value perspectives. In this study, emphasized public values were accessibility, citizen engagement, transparency, responsiveness, dialog and balancing of interests. Akıncılar and Dağdeviren (2014) proposed a hybrid model including Analytic Hierarchy Process (AHP) and Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) to evaluate websites of five-star hotels in Ankara. The AHP was utilized to weigh criteria and, ranking of alternatives were produced by PROMETHEE. Cao and Yang (2016) evaluated Chinese tourism websites using a content analysis technique. This technique assessed website performances in term of e-commerce adoption and revealed significant differences in performance between the website types. As a result, online travel agencies were found to perform better than other types of tourism websites. Kang, Jang and Park (2016) presented a fuzzy hierarchical TOPSIS based on E-S-QUAL (the extended version of SERVQUAL) for measurement of electronic service quality in the e-commerce environment. In this study, website evaluation criteria were separated in four dimensions and 22 sub-criteria were defined in these four dimensions. E-S-QUAL was used to determine criteria weights. Fuzzy hierarchical TOPSIS was implemented to rank e-commerce websites. Díaz, et al. (2017) presented a set of 12 cultural-oriented usability heuristics to evaluate e-Commerce websites. İlbahar and Cebi (2017) used fuzzy KANO model and classified design parameters according to customer expectations in order to evaluate the usability of e-commerce websites. Four websites were ranked respect to the usability scores. Chen (2018) proposed a mathematical programming model for usability evaluation. Additionally, the model was tested on a large synthetic dataset to demonstrate its remarkable scalability. This model reduced information load by removing links from highly clustered pages while minimizing the impact for users. Results indicate that the model not only significantly reduces page complexity with little impact on user navigation, but also can be solved effectively. Chmielarza and Zborowskib (2018) identified the best e-banking websites using TOPSIS in Poland from the view point of individual clients. Three main criteria group as economic criteria, technical, visual and security criteria, criteria related to anti-crisis measures and 20 sub-criteria were determined in these three group. In order to obtain the data for analysis, an electronic spreadsheet of evaluations was created on the Internet and participants were selected with purposeful random sampling. Subsequently, on the basis of the obtained scores,

authors carried out analyses and presented findings. Verkijika and Wet (2018) evaluated a total of 279 e-government websites from 31 Sub-Saharan Africa (SSA) countries and found that most e-government websites in SSA have poor usability. As a result of the study, usability of e-government websites was positively associated with the E-Government Development Index and the E-Participation Index. Pamucar et al. (2018) supposed a hybrid Interval Rough Analytic Hierarchy Process-Multi Attributive Border Approximation Area Comparison (IR-AHP-MABAC) model for evaluating quality of university websites. The model can treat uncertainties through the use of interval rough numbers (IRN). The modified IR-AHP method was used to determine criteria weight coefficients while IR-MABAC model was applied for comparing university websites. Lee-Geilera (2019) developed Democratic E-governance Website Evaluation Model as an integrative model for evaluating a government website. This model consists of three dimensions as transparency, service quality, and citizen engagement, based on the democratic e-governance framework. Additionally, a qualitative meta-analysis of four literatures including information systems, business administration, public administration, and democratic theory was conducted in the study. Ostovare and Shahraki (2019) evaluated status of websites and e-services provided by five-star hotels. Initial criteria and sub-criteria were determined with fuzzy Delphi method. In addition, Shannon entropy was used to determine the criteria weights. PROMETHEE and Geometrical Analysis for Interactive Aid (GAIA) methods were used to rank and improve the visual aids of websites. As a result of the study, it was determined that customer focus is the most important criterion, followed by marketing, security and technology. Liang et al. (2019) provided a new perspective for a compromised solution, which can handle the decision maker's psychological behavior by inducing TODIM (a Portuguese acronym meaning Interactive Multi-Criteria Decision Making). In the study, Pythagorean fuzzy entropy and cross-entropy measures were used to compute criteria weights. Then, considering the psychological behavior of the customer, two types of strategies for the combination TODIM and VIKOR were suggested. After that, a simulated example of ranking internet banking websites in the Ghanaian banking industry was given to illustrate the validity and applicability of the proposed two strategies. Finally, Wilcoxon signed-rank test was performed to discuss differences between VIKOR, TODIM and the proposed two strategies.

In terms of evaluating websites usability of airline firms, there are limited study in the literature. However, airline firms aim to increase their business value and make their websites more usable to facilitate e-ticketing or online transactions. In the air transport industry, web-based marketing is widely applied to attract new customers as well as serving customers continuously. Gutierrez et al. (2005) investigated the disability-accessibility of the online reservations systems of

commercial US airline carriers, as well as, those of foreign carriers that fly to/from the US. Lituchy and Barra (2008) investigated the effects of international differences on website design, implementation and usage. In this study, the websites of airlines and hotels were examined and it was observed that the websites consider the language and cultural differences. Important recommendations have been made for improvement for both sectors. Harison and Boonstra (2008) examined methods used in the assessment of online operations and AFWSs. This model was used to measure the commercial, technical, and operational aspects of websites. Tsai et al. (2011) proposed an integrated model for marketing mix 4P and website quality to evaluate the effectiveness of AFWSs and this model was applied to analyze the websites of five air transportation companies in Taiwan. In this model, DEMATEL was used to analyze the relationship between criteria, while ANP was applied to determine the criteria weights. VIKOR was performed to rank the performance of websites. Díaz and Consuegra (2016) applied the latent class segmentation model to identify different clusters of airlines based on the website evaluation. This study has enabled the identification and definition of significant differences in the degree of persuasiveness displayed by the websites. The model includes six dimensions of website persuasiveness (informativeness, usability, credibility, inspiration, involvement and reciprocity) and it was applied based on content analysis of a sample of 240 airline websites. Three airline segments identified as “influential”, “follower”, and “passive” and the results were provided valuable information for airlines’ managers to understand the relative competitiveness level of their websites in the airline industry.

As seen from the literature, user test and survey results have not been evaluated at the same time by performing MCDM approaches for AFWSs’ usability evaluations. Additionally, MCDM methods implemented in the usability evaluation are the classical and commonly used methods which used for the other decision problems. In this manner, the proposed MDUEA has many originalities in terms of combined methodologies, considered usability criteria, evaluating AFWSs’ usability taking into account both measurable data and users’ opinions. This study is the first in the literature that evaluates AFWSs’ usability in such detail for more than one firms at the same time.

3. THE PROPOSED MULTI-DIMENSIONAL USABILITY EVALUATION APPROACH

The proposed MDUEA includes four phases as *determining components of usability assessment, usability testing with MORAE V3, computing usability criteria weights by*

implementing different weighting approaches for objective and subjective usability criteria and aggregating weight results of these approaches, ranking webpages by performing CODAS method. The flowchart of the proposed approach is given in Figure 1.

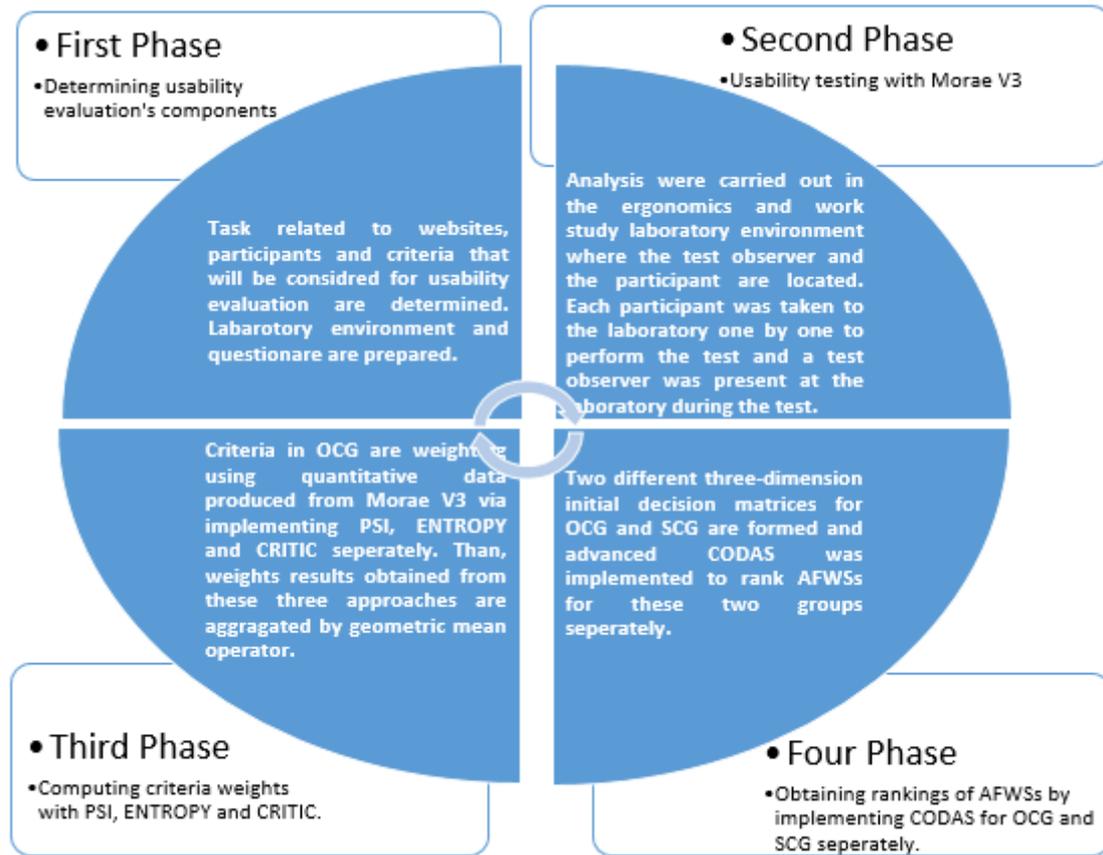


Figure 1. The flow chart of the proposed approach

The main components used in the study are given in Table 1.

Table 1. Main Components

Notation	Definition	Application
$T_i, i = 1, \dots, I$	Task	$T_i, i = 1, \dots, 5$ T_1 -Selecting departure and arrival flights and purchasing tickets T_2 -Online check in T_3 -Seat selection T_4 -Baggage tracking T_5 -Learning flight status
$P_j, j = 1, \dots, J$	Participants	$P_j, j = 1, \dots, 40$ 40 students having predetermined features among 70 students are selected.
$OC_l, l = 1, \dots, L$	Objective criteria	$OC_l, l = 1, \dots, 5$ OC_1 -Task time OC_2 -Time between two data entries OC_3 -Number of mouse clicks

			OC_4 -Mouse movement distance
			OC_5 -Success percentage
			SC_1 -Level of willingness to use
			SC_2 -Level of complexity
			SC_3 -Easy to use level
			SC_4 -Level of need for technical support
			SC_5 -Feeling safe while using
$SC_m, m = 1, \dots, M$	Subjective criteria	$SC_m, m = 1, \dots, 10$	SC_6 -Level of required prior knowledge
			SC_7 -Level of integration goodness
			SC_8 -Level of inconsistency
			SC_9 -Learning speed
			SC_{10} -Cumbersome level
$WS_V, v = 1, \dots, V$	Website	$WS_V, v = 1, \dots, 5$	The websites' usability of the five most preferred airline firm in Turkey were evaluated in term of domestic flights

1.1. Participants, laboratory environment and questionnaire

70 fourth-class industrial engineering students were participated in the study and they evaluated the usability of the AFWSs. The reasons of selection the fourth-class students are the number of lessons they take is less than the other classes and therefore they have more opportunity to go to their families. The sample was made up of senior students whose family lived outside Ankara. In addition, the answers of the students who prefer the airline most by means of transportation were considered. Again, the evaluations of those students who go to their families with an airline firm at least once every few months were considered for usability evaluation of AFWSs. Since, 40 participants among 70 students have these features, the participation rate was determined as 57%. Demographic information about the related students are given in Table 2.

Table 2. Demographic information for students

Demographic features		Frequency for 70	%	Frequency for 40	%
Gender	Male	27	38.60	15	37.50
	Female	43	61.40	25	62.50
Age	20	12	17.14	5	12.50
	21	20	28.57	15	37.50
	22	11	15.71	9	22.50
	23	14	20.00	5	12.50
	24 +	13	18.57	6	15.00

Family income status	Low	16	22.86	10	25.00
	Medium	11	15.71	1	2.50
	High	20	28.57	14	35.00
	Very high	23	32.86	15	37.50
Frequency of airline usage	I don't use airline while traveling.	8	11.43	0	0.00
	once a week	14	20.00	14	35.00
	once a month	18	25.71	18	45.00
	every few months	8	11.43	8	20.00
	One a year	10	14.29	0	0.00
	Other	12	17.14	0	0.00
The most served airline company	Airline firm 1	8	11.43	0	0.00
	Airline firm 2	7	10.00	0	0.00
	Airline firm 3	9	12.86	0	0.00
	Airline firm 4	3	4.29	0	0.00
	Airline firm 5	3	4.29	0	0.00
	All	40	57.14	40	100.00
The most preferred travel way for transportation	Airplane	40	57.14	40	100.00
	Bus	13	18.57	0	0.00
	Train	9	12.86	0	0.00
	private car	8	11.43	0	0.00

According to Nielsen (1993), 5 participants are sufficient to test systems, and these participants can detect existing errors by 75%. When the number of subjects increases, the number of detected errors does not show a remarkable increase. In order to obtain results at approximately 90% reliability level; 13 users should evaluate the related system (Nielsen, 1993).

In the study, usability of five most preferred airline firms' web pages in Turkey were evaluated in term of domestic flights. Airline firms were chosen by a focus group including two managers working in the airline industry and two academicians who are experts in the marketing and passenger transportation disciplines. The first and second firms serve international and national passengers based on a full-service carrier strategy within a well-known business network, while the third, fourth and fifth built their business strategies upon low cost carrier strategies. Among these firms, third one does not provide any additional services, such as in-flight catering and frequent traveler lounge. Fourth offers paid in-flight and on-ground services.

In the study, a survey consisting of three parts was applied. The first part of the survey includes questions for determining the demographic characteristics and travel preferences of participants. In the second part, SUS is located. In the third part, the participants were asked to rank the criteria in the SCG from the most important to the least important, from 1 to 10.

Questions within the scope of SUS were asked to the participants for the evaluation of the five AFWSs after performing predetermined tasks. SUS considers some phrases related to the usability of web pages and it uses 1-5 scale (1: Strongly disagree and 5: Strongly agree). The SUS is a simple, ten-item scale giving a global view of subjective assessments of usability. SUS covers a variety of aspects of system usability, such as the need for support, training, and complexity, and thus have a high level of face validity for measuring usability of a system. The SUS is shown in the appendix (Brooke, 1996).

First phase: Determining components of usability assessment.

In this phase, components of usability evaluation as tasks, participants, usability criteria are determined. Tasks are required to test the usability of websites and tasks are performed by end users. End users are formed participants. Usability criteria are performance measurements for the websites and these criteria are divided into two groups as OCG and SCG. The performance values of websites for OCG were obtained by applying MORAE V3 and SCG was obtained by rankings criteria in SCG from the most important to the least important by participants. Websites are alternatives compared in terms of their usability. In this study, five AFWSs were ordered for their usability levels. To compare these websites' usability, 40 students were chosen considering some features in the study. The tasks were determined by taking into account the most common operations performed by students related to the AFWSs for domestic flights by the researchers of the study. According to Nielsen (1993), the basic rule when deciding test tasks for the software's usability assessment is that it represents as much of the users' tasks as possible (Nielsen, 1993).

Second phase: Usability testing with MORAE V3

Within the scope of the analysis, one computer (Intel i7 processor, 8 GB RAM, Windows 10 operating system), one microphone, Morae V3 2.1. usability test program, Microsoft Edge Internet browser were used. The ergonomics and work study laboratory of the engineering faculty of a university in Ankara, Turkey was used for analysis. Analysis were carried out in the laboratory environment where the test observer and the participant are located. Each participant was taken to the laboratory one by one to perform the test and a test observer was present at the laboratory during the test.

The Morae V3 package program used for the tests can record the user actions (user face, sound, screen, mouse movements); it can automatically provide standard measurement data (such as mouse clicks, mouse movement distance, task time). The SUS is performed by Morae software at the end of the task. Before starting the analysis, the participants were informed about the

purpose and content of the test. After these meetings, the test was started with the Morae program. During the test, the tasks are seen one by one on the screen as a window. Each participant completed the SUS after completing five different tasks for each website. Each participant performed 5 tasks for each website every 3 days. 200 tests have been implemented for all websites totally during 3 months.

Third phase: Computing usability criteria weights by implementing different weighting approaches and aggregating the weight results of these approaches.

To perform objective weighting approaches, three-dimension initial decision matrix $[X]$ depicted in Table 3 should be structured firstly. $[X]$ includes performance values of each website for each criterion in terms of tasks. Each element of $[X]$ denoted as $x_{v_i_l}$ was obtained by averaging $x_{j_{v_i_l}}$ values. $x_{v_i_l}$ is obtained by averaging $x_{p_{v_i_l}}$ values and $x_{j_{v_i_l}}$ is l th criterion value in OCG for the j th participant related to the v th web page when performing i th task. $[X]$ is used for all weighting approach in the same way.

Step 3.1. Computing criteria weights in OCG.

To compute each criterion weight in OCG, three different weighting approaches as CRITIC, PSI and ENTROPY are implemented.

Step 3.1.1. Computing objective usability criteria weights by implementing PSI.

In PSI, criteria are divided into two groups as cost and benefit. Among criteria in OCG, excluding OC_5 , the other criteria are cost type. After structuring $[X]$, this matrix should be normalized because of including values with different units and normalized initial decision matrix $[N]$ is obtained (Table 4). Since, $[X]$ includes cost and benefit type criteria, two different normalization processes are performed in PSI as in Eq.(1) and Eq.(2). For example, $x_{1_{1_1}}$ is average task time (OC_1) of all 40 participants for the first web page (WP_1) when they perform the first task T_1 . $x_{1_{1_1}}$ is the task time (OC_1) of the first participant (P_1) when performing the first task (T_1) on the first web page (WP_1).

For benefit type criteria in OCG,

$$n_{v_i_l} = \frac{x_{v_i_l}}{x_{i_l_{max}}} \quad (1)$$

For cost type criteria in OCG,

$$n_{v_i_l} = \frac{x_{i_l_{min}}}{x_{v_i_l}} \quad (2)$$

Here; n_{v_i} is the normalized value of lth criterion in OCG for the vth web page when performing ith task. n_{1_1} is normalized task time (OC_1) of all 40 participants for the first web page (WP_1) when they perform the first task T_1

$x_{i_{lmax}}$ is the maximum value of lth criterion among all websites when performing ith task and $x_{i_{lmin}}$ is the minimum value of lth criterion among all websites when performing ith task. The next step is to compute preference variation for each criterion in OCG for each task denoted as PV_{i_l} as in Eq.(3).

$$PV_{i_l} = \sum_{v=1}^V [n_{v_i} - \bar{n}_{i_l}]^2 \quad (3)$$

Here, \bar{n}_{i_l} is the average value of n_{v_i} and it is computed as $\bar{n}_{i_l} = \frac{1}{V} \sum_{v=1}^V n_{v_i}$. PV_{i_l} values are given in Table 5 for PSI. The last step of the PSI is to calculate deviation in preference value for each criterion for each task denoted as DPV_{i_l} by using Eq.(4). DPV_{i_l} values are given in Table 6.

$$DPV_{i_l} = 1 - PV_{i_l} \quad (4)$$

Steps up to here create the traditional procedure of PSI method. Steps performed after that are procedures developed by the authors. To obtain the criteria weights, DPV_{i_l} value of each criterion for all tasks were summed and, in this way, performance of websites for each task based on each criterion denoted as DPV_{l_t} were obtained as in Eq.(5).

$$DPV_{l_t} = \sum_{i=1}^I DPV_{i_l} \quad (5)$$

Then, all DPV_{l_t} values are normalized to obtain each criterion weight ($w_{l_{PSI}}$) between [0, 1] as in Eq.(6) based on PSI. $w_{l_{PSI}}$ values are depicted in Table 7.

$$w_{l_{PSI}} = \frac{DPV_{l_t}}{\sum_{l=1}^L DPV_{l_t}}, \sum_{l=1}^L w_{l_{PSI}} = 1 \quad (6)$$

Table 3. Three-dimension initial decision matrix for OCG

WP_V	T_1					T_2					T_3					T_4					T_5				
	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5
	$x_{v_{11}}$	$x_{v_{12}}$	$x_{v_{13}}$	$x_{v_{14}}$	$x_{v_{15}}$	$x_{v_{21}}$	$x_{v_{22}}$	$x_{v_{23}}$	$x_{v_{24}}$	$x_{v_{25}}$	$x_{v_{31}}$	$x_{v_{32}}$	$x_{v_{33}}$	$x_{v_{34}}$	$x_{v_{35}}$	$x_{v_{41}}$	$x_{v_{42}}$	$x_{v_{43}}$	$x_{v_{44}}$	$x_{v_{45}}$	$x_{v_{51}}$	$x_{v_{52}}$	$x_{v_{53}}$	$x_{v_{54}}$	$x_{v_{55}}$
WP_1	37,19	7,59	9,78	11033,43	81,10	22,73	6,24	7,27	6464,69	81,29	34,88	5,78	9,59	13425,34	81,59	31,80	8,19	7,29	7681,06	81,90	32,07	6,82	6,93	7654,48	84,76
WP_2	36,20	11,81	12,54	13927,15	78,66	37,57	12,93	10,54	7518,25	79,73	37,05	10,81	9,95	10061,56	81,34	43,66	11,94	10,90	10491,79	80,24	36,75	12,65	11,00	9568,21	81,34
WP_3	36,34	12,38	11,15	15730,38	80,83	32,97	15,66	12,29	11535,33	78,39	36,38	10,97	10,83	12192,90	81,78	34,68	12,20	10,85	10138,18	83,07	40,49	12,22	11,15	9319,42	78,95
WP_4	35,53	13,87	12,17	19317,71	79,17	37,04	37,04	10,22	10169,89	78,85	39,11	12,09	9,93	9184,96	82,66	41,31	9,01	9,93	8625,32	80,24	39,79	11,69	11,37	10105,43	79,41
WP_5	37,67	14,81	12,98	16030,26	79,15	35,61	14,19	11,32	10517,40	79,15	38,41	11,45	11,22	11579,12	81,51	39,25	11,61	11,61	7700,11	81,22	34,94	11,42	11,42	6514,05	82,29

Table 4. Normalized initial decision matrix for PSI

WP_V	T_1					T_2					T_3					T_4					T_5				
	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5
	$n_{v_{11}}$	$n_{v_{12}}$	$n_{v_{13}}$	$n_{v_{14}}$	$n_{v_{15}}$	$n_{v_{21}}$	$n_{v_{22}}$	$n_{v_{23}}$	$n_{v_{24}}$	$n_{v_{25}}$	$n_{v_{31}}$	$n_{v_{32}}$	$n_{v_{33}}$	$n_{v_{34}}$	$n_{v_{35}}$	$n_{v_{41}}$	$n_{v_{42}}$	$n_{v_{43}}$	$n_{v_{44}}$	$n_{v_{45}}$	$n_{v_{51}}$	$n_{v_{52}}$	$n_{v_{53}}$	$n_{v_{54}}$	$n_{v_{55}}$
WP_1	0,96	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,68	0,99	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,85	1,00
WP_2	0,98	0,64	0,78	0,79	0,97	0,60	0,48	0,69	0,86	0,98	0,94	0,53	0,96	0,91	0,98	0,73	0,69	0,67	0,73	0,97	0,87	0,54	0,63	0,68	0,96
WP_3	0,98	0,61	0,88	0,70	1,00	0,69	0,40	0,59	0,56	0,96	0,96	0,53	0,89	0,75	0,99	0,92	0,67	0,67	0,76	1,00	0,79	0,56	0,62	0,70	0,93
WP_4	1,00	0,55	0,80	0,57	0,98	0,61	0,17	0,71	0,64	0,97	0,89	0,48	0,97	1,00	1,00	0,77	0,91	0,73	0,89	0,97	0,81	0,58	0,61	0,64	0,94
WP_5	0,94	0,51	0,75	0,69	0,98	0,64	0,44	0,64	0,61	0,97	0,91	0,50	0,85	0,79	0,99	0,81	0,71	0,63	1,00	0,98	0,92	0,60	0,61	1,00	0,97

Table 5. Preference variation value for each criterion in OCG

T_1					T_2					T_3					T_4					T_5				
OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5
PV_{11}	PV_{12}	PV_{13}	PV_{14}	PV_{15}	PV_{21}	PV_{22}	PV_{23}	PV_{24}	PV_{25}	PV_{31}	PV_{32}	PV_{33}	PV_{34}	PV_{35}	PV_{41}	PV_{42}	PV_{43}	PV_{44}	PV_{45}	PV_{51}	PV_{52}	PV_{53}	PV_{54}	PV_{55}
0,002	0,1527	0,0393	0,1024	0,0007	0,1100	0,3741	0,1018	0,1406	0,0008	0,0073	0,1932	0,0149	0,0643	0,0002	0,0496	0,0900	0,0899	0,0651	0,0008	0,0290	0,1504	0,1179	0,0881	0,0031

Table 6. Deviation in preference value for each criterion in OCG

T_1					T_2					T_3					T_4					T_5				
OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5
DPV_{11}	DPV_{12}	DPV_{13}	DPV_{14}	DPV_{15}	DPV_{21}	DPV_{22}	DPV_{23}	DPV_{24}	DPV_{25}	DPV_{31}	DPV_{32}	DPV_{33}	DPV_{34}	DPV_{35}	DPV_{41}	DPV_{42}	DPV_{43}	DPV_{44}	DPV_{45}	DPV_{51}	DPV_{52}	DPV_{53}	DPV_{54}	DPV_{55}
0,9980	0,8473	0,9607	0,8976	0,9993	0,8900	0,6259	0,8982	0,8594	0,9992	0,9927	0,8068	0,9851	0,9357	0,9998	0,9504	0,9100	0,9101	0,9349	0,9992	0,9710	0,8496	0,8821	0,9119	0,9969

Table 7. Criteria weights in OCG for PSI

OC_1	OC_2	OC_3	OC_4	OC_5
w_{1PSI}	w_{2PSI}	w_{3PSI}	w_{4PSI}	w_{5PSI}
0.17	0.18	0.21	0.21	0.23

Step 3.1.2. Computing objective usability criteria weights by implementing ENTROPY

ENTROPY was another method used to obtain criteria weights in OCG. Again, $[X]$ (Table 3) was used to obtain $[N]$ as in Eqs.(1) and (2). $n_{v_{i_l}}$ values for ENTROPY are given in Table 8.

$$n_{v_{i_l}} = \frac{x_{v_{i_l}}}{\sum_{v=1}^V x_{v_{i_l}}} \quad (7)$$

Then, entropy value for each criterion for each task denoted as e_{i_l} were computed via utilizing Eq.(8).

$$e_{i_l} = -k \sum_{v=1}^V n_{v_{i_l}} \ln(n_{v_{i_l}}) \quad (8)$$

Here, k is the entropy coefficient and it is calculated as $\frac{1}{\ln(L)}$. According to e_{i_l} given in Table 9, the amount of entropy removal values (d_{i_l}) depicted in Table 10 are obtained as in Eq.(9).

$$d_{i_l} = 1 - e_{i_l} \quad (9)$$

Steps up to here create the traditional procedure of ENTROPY method. Steps performed after that are procedures developed by the authors. To obtain criteria weights, d_{i_l} value of each criterion for each task were summed and d_t were obtained as in Eq.(10). d_t values are given in Table 9.

$$d_{l_t} = \sum_{i=1}^I d_{i_l} \quad (10)$$

Then, all d_{l_t} values are normalized to obtain each criterion weight ($w_{l_{ENT}}$) between $[0, 1]$ as in Eq.(11). $w_{l_{ENT}}$ values are depicted in Table 11.

$$w_{l_{ENT}} = \frac{d_{l_t}}{\sum_{l=1}^L d_{l_t}}, \sum_{l=1}^L w_{l_{ENT}} = 1 \quad (11)$$

Table 8. Normalized initial decision matrix for ENTROPY

WP_V	T_1					T_2					T_3					T_4					T_5				
	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5
	$n_{v_{11}}$	$n_{v_{12}}$	$n_{v_{13}}$	$n_{v_{14}}$	$n_{v_{15}}$	$n_{v_{21}}$	$n_{v_{22}}$	$n_{v_{23}}$	$n_{v_{24}}$	$n_{v_{25}}$	$n_{v_{31}}$	$n_{v_{32}}$	$n_{v_{33}}$	$n_{v_{34}}$	$n_{v_{35}}$	$n_{v_{41}}$	$n_{v_{42}}$	$n_{v_{43}}$	$n_{v_{44}}$	$n_{v_{45}}$	$n_{v_{51}}$	$n_{v_{52}}$	$n_{v_{53}}$	$n_{v_{54}}$	$n_{v_{55}}$
WP_1	0,20	0,13	0,17	0,15	0,20	0,14	0,07	0,14	0,14	0,20	0,19	0,11	0,19	0,24	0,20	0,17	0,15	0,14	0,17	0,20	0,17	0,12	0,13	0,18	0,21
WP_2	0,20	0,20	0,21	0,18	0,20	0,23	0,15	0,20	0,16	0,20	0,20	0,21	0,19	0,18	0,20	0,23	0,23	0,22	0,24	0,20	0,20	0,23	0,21	0,22	0,20
WP_3	0,20	0,20	0,19	0,21	0,20	0,20	0,18	0,24	0,25	0,20	0,20	0,21	0,21	0,22	0,20	0,18	0,23	0,21	0,23	0,20	0,22	0,22	0,21	0,22	0,19
WP_4	0,19	0,23	0,21	0,25	0,20	0,22	0,43	0,20	0,22	0,20	0,21	0,24	0,19	0,16	0,20	0,22	0,17	0,20	0,19	0,20	0,22	0,21	0,22	0,23	0,20
WP_5	0,21	0,24	0,22	0,21	0,20	0,21	0,16	0,22	0,23	0,20	0,21	0,22	0,22	0,21	0,20	0,21	0,22	0,23	0,17	0,20	0,19	0,21	0,22	0,15	0,20

Table 9. Entropy value for each criterion in OCG

T_1					T_2					T_3					T_4					T_5				
OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5
e_{1_1}	e_{1_2}	e_{1_3}	e_{1_4}	e_{1_5}	e_{2_1}	e_{2_2}	e_{2_3}	e_{2_4}	e_{2_5}	e_{3_1}	e_{3_2}	e_{3_3}	e_{3_4}	e_{3_5}	e_{4_1}	e_{4_2}	e_{4_3}	e_{4_4}	e_{4_5}	e_{5_1}	e_{5_2}	e_{5_3}	e_{5_4}	e_{5_5}
1,000	0,986	0,997	0,990	1,000	0,991	0,898	0,991	0,986	1,000	0,999	0,983	0,999	0,994	1,000	0,996	0,992	0,993	0,995	1,000	0,998	0,987	0,990	0,992	1,000

Table 10. The amount of entropy removal value for each criterion in OCG

T_1					T_2					T_3					T_4					T_5					d_t
OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	
d_{1_1}	d_{1_2}	d_{1_3}	d_{1_4}	d_{1_5}	d_{2_1}	d_{2_2}	d_{2_3}	d_{2_4}	d_{2_5}	d_{3_1}	d_{3_2}	d_{3_3}	d_{3_4}	d_{3_5}	d_{4_1}	d_{4_2}	d_{4_3}	d_{4_4}	d_{4_5}	d_{5_1}	d_{5_2}	d_{5_3}	d_{5_4}	d_{5_5}	
0,000	0,014	0,003	0,010	0,000	0,009	0,102	0,009	0,014	0,000	0,001	0,017	0,001	0,006	0,000	0,004	0,008	0,007	0,005	0,000	0,002	0,013	0,010	0,008	0,000	0,243

Table 11. Weights of objective criteria for ENTROPY

OC_1	OC_2	OC_3	OC_4	OC_5
W_{1ENT}	W_{2ENT}	W_{3ENT}	W_{4ENT}	W_{5ENT}
0.066	0.634	0.123	0.176	0.002

Step 3.1.3. Computing objective usability criteria weights by implementing CRITIC.

CRITIC is the last method used to obtain criteria weights in OCG. Again, [X] (Table 3) is used to obtain [N] given in Table 12 by considering cost and benefit type criteria separately as in Eq.(12) and Eq.(13).

For benefit type criteria in OCG,

$$n_{v_{i_l}} = \frac{x_{v_{i_l}} - x_{i_{l_{min}}}}{x_{i_{l_{max}}} - x_{i_{l_{min}}}} \quad (12)$$

For cost type criteria in OCG,

$$n_{v_{i_l}} = \frac{x_{i_{l_{max}}} - x_{v_{i_l}}}{x_{i_{l_{max}}} - x_{i_{l_{min}}}} \quad (13)$$

Then, standard deviation for each criterion (σ_{i_l}) for each task for each website depicted in Table 12 is obtained according to $n_{v_{i_l}}$ as in Eq.(14).

$$\sigma_{i_l} = \sqrt{\frac{\sum_{v=1}^V (n_{v_{i_l}} - \bar{n}_{i_l})^2}{V}} \quad (14)$$

Here, \bar{n}_{i_l} is the mean value of $n_{v_{i_l}}$ for each task for each criterion. After obtaining σ_{i_l} , average standard deviation ($\bar{\sigma}_l$) given in Table 13 was computed for each criterion as in Eq.(15).

$$\bar{\sigma}_l = \frac{\sum_{i=1}^I \sigma_{i_l}}{L} \quad (15)$$

CRITIC considers relationship between criteria when determining criteria weights. Correlation between criteria in OCG for each task seen in Table 14 were examined by implementing Spearman Correlation Coefficient ($\rho_{i_{lk}}$) as in Eq.(16) for each task.

$$\rho_{i_{lk}} = 1 - \frac{6 \sum_{l=1}^L d_{il}}{L(L^2 - 1)}; k = 1, \dots, L \quad (16)$$

According to computed $\bar{\sigma}_l$ and $\rho_{i_{lk}}$ values, the amount of information obtained for each criterion for each task (S_{i_l}) given in Table 15 is determined as in Eq.(17).

$$S_{i_l} = 1 - \rho_{i_{lk}} \quad (17)$$

Table 12. Normalized Initial decision matrix and standard deviation of each criterion for CRITIC

WP_V	T_1					T_2					T_3					T_4					T_5					
	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	
	$n_{v_{11}}$	$n_{v_{12}}$	$n_{v_{13}}$	$n_{v_{14}}$	$n_{v_{15}}$	$n_{v_{21}}$	$n_{v_{22}}$	$n_{v_{23}}$	$n_{v_{24}}$	$n_{v_{25}}$	$n_{v_{31}}$	$n_{v_{32}}$	$n_{v_{33}}$	$n_{v_{34}}$	$n_{v_{35}}$	$n_{v_{41}}$	$n_{v_{42}}$	$n_{v_{43}}$	$n_{v_{44}}$	$n_{v_{45}}$	$n_{v_{51}}$	$n_{v_{52}}$	$n_{v_{53}}$	$n_{v_{54}}$	$n_{v_{55}}$	
WP_1	0,96	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,68	0,99	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
WP_2	0,98	0,64	0,78	0,79	0,97	0,60	0,48	0,69	0,86	0,98	0,94	0,53	0,96	0,91	0,98	0,73	0,69	0,67	0,73	0,97	0,87	0,54	0,63	0,80	0,96	
WP_3	0,98	0,61	0,88	0,70	1,00	0,69	0,40	0,59	0,56	0,96	0,96	0,53	0,89	0,75	0,99	0,92	0,67	0,67	0,76	1,00	0,79	0,56	0,62	0,82	0,93	
WP_4	1,00	0,55	0,80	0,57	0,98	0,61	0,17	0,71	0,64	0,97	0,89	0,48	0,97	1,00	1,00	0,77	0,91	0,73	0,89	0,97	0,81	0,58	0,61	0,76	0,94	
WP_5	0,94	0,51	0,75	0,69	0,98	0,64	0,44	0,64	0,61	0,97	0,91	0,50	0,85	0,79	0,99	0,81	0,71	0,63	1,00	0,98	0,92	0,60	0,61	1,18	0,97	
σ_i	0,02	0,17	0,09	0,14	0,01	0,15	0,27	0,14	0,17	0,01	0,04	0,20	0,05	0,11	0,01	0,10	0,13	0,13	0,11	0,01	0,08	0,17	0,15	0,16	0,02	

Table 13. Average standard deviation values for each criterion in OCG

OC_1	OC_2	OC_3	OC_4	OC_5
$\bar{\sigma}_1$	$\bar{\sigma}_1$	$\bar{\sigma}_1$	$\bar{\sigma}_1$	$\bar{\sigma}_1$
0,076	0,190	0,115	0,139	0,014

Table 14. Spearman correlation coefficient for criterion for each task

OC_i	T_1					T_2					T_3					T_4					T_5				
	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5
	ρ_{11k}	ρ_{12k}	ρ_{13k}	ρ_{14k}	ρ_{15k}	ρ_{21k}	ρ_{22k}	ρ_{23k}	ρ_{24k}	ρ_{25k}	ρ_{31k}	ρ_{32k}	ρ_{33k}	ρ_{34k}	ρ_{35k}	ρ_{41k}	ρ_{42k}	ρ_{43k}	ρ_{44k}	ρ_{45k}	ρ_{51k}	ρ_{52k}	ρ_{53k}	ρ_{54k}	ρ_{55k}
OC_1	1,000	0,051	0,205	-0,359	-0,189	1,000	0,300	0,100	0,000	0,051	1,000	0,975	0,300	0,900	-0,447	1,000	0,300	0,359	0,616	0,872	1,000	0,700	0,410	0,600	1,000
OC_2	0,051	1,000	0,700	0,900	0,264	0,300	1,000	0,400	0,700	0,821	0,975	1,000	0,359	-0,821	-0,574	0,300	1,000	0,667	0,718	-0,205	0,700	1,000	0,051	0,600	0,700
OC_3	0,205	0,700	1,000	0,500	0,791	0,100	0,400	1,000	0,900	0,821	0,300	0,359	1,000	-0,100	0,224	0,359	0,667	1,000	0,158	0,026	0,410	0,051	1,000	0,103	0,410
OC_4	-0,359	0,900	0,500	1,000	0,264	0,000	0,700	0,900	1,000	0,975	-0,900	-0,821	-0,100	1,000	0,224	0,616	0,718	0,158	1,000	0,263	0,600	0,600	0,103	1,000	0,600
OC_5	-0,189	0,264	0,791	0,264	1,000	0,051	0,821	0,821	0,975	1,000	-0,447	-0,574	0,224	0,224	1,000	0,872	-0,205	0,026	0,263	1,000	1,000	0,700	0,410	0,600	1,000

Table 15. The amount of information obtained for each criterion for each task and the total amount of information for each criterion

OC_i	T_1					T_2					T_3					T_4					T_5				
	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5	OC_1	OC_2	OC_3	OC_4	OC_5
	S_{11k}	S_{12k}	S_{13k}	S_{14k}	S_{15k}	S_{21k}	S_{22k}	S_{23k}	S_{24k}	S_{25k}	S_{31k}	S_{32k}	S_{33k}	S_{34k}	S_{35k}	S_{41k}	S_{42k}	S_{43k}	S_{44k}	S_{45k}	S_{51k}	S_{52k}	S_{53k}	S_{54k}	S_{55k}
OC_1	0,000	0,949	0,795	1,359	1,189	0,000	0,700	0,900	1,000	0,949	0,000	0,025	0,700	0,100	1,447	0,000	0,700	0,641	0,384	0,128	0,000	0,300	0,590	0,400	0,000
OC_2	0,949	0,000	0,300	0,100	0,736	0,700	0,000	0,600	0,300	0,179	0,025	0,000	0,641	1,821	1,574	0,700	0,000	0,333	0,282	1,205	0,300	0,000	0,949	0,400	0,300
OC_3	0,795	0,300	0,000	0,500	0,209	0,900	0,600	0,000	0,100	0,179	0,700	0,641	0,000	1,100	0,776	0,641	0,333	0,000	0,842	0,974	0,590	0,949	0,000	0,897	0,590
OC_4	1,359	0,100	0,500	0,000	0,736	1,000	0,300	0,100	0,000	0,025	1,900	1,821	1,100	0,000	0,776	0,384	0,282	0,842	0,000	0,737	0,400	0,400	0,897	0,000	0,400
OC_5	1,189	0,736	0,209	0,736	0,000	0,949	0,179	0,179	0,025	0,000	1,447	1,574	0,776	0,776	0,000	0,128	1,205	0,974	0,737	0,000	0,000	0,300	0,590	0,400	0,000
$\sum_{i=1}^I S_{iik}$	4,292	2,085	1,804	2,695	2,870	3,549	1,779	1,779	1,425	1,332	4,072	4,061	3,217	3,797	4,573	2,520	2,790	2,245	3,044	2,520	1,290	1,949	3,026	2,097	1,290

Then, S_{i_l} values in Table 15 are summed for each criterion to compute the total information amount for each criterion (S_l) as in Table 16 as in Eq.(18).

$$S_l = \sum_{l=1}^L \sum_{k=1}^L S_{lk} \quad (18)$$

Table 16. Total information amount for each criterion

OC_1	OC_2	OC_3	OC_4	OC_5
S_1	S_2	S_3	S_4	S_5
15.056	12.394	12.616	12.259	13.109

Finally, to compute criteria weights ($w_{l_{CRT}}$) as in Table 17, Eq.(19) is used.

$$w_{l_{CRT}} = \frac{S_l \times \bar{\sigma}_l}{\sum_{l=1}^L S_l \times \bar{\sigma}_l}, \sum_{l=1}^L w_{l_{CRT}} = 1.00 \quad (19)$$

Table 17. Weights of objective criteria for CRITIC

OC_1	OC_2	OC_3	OC_4	OC_5
$w_{1_{CRT}}$	$w_{2_{CRT}}$	$w_{3_{CRT}}$	$w_{4_{CRT}}$	$w_{5_{CRT}}$
0.168	0.345	0.212	0.249	0.026

Step 3.1.4. Computing aggregated criteria weights in OCG.

To aggregate weights of criteria in OCG obtained from PSI, ENTROPY and CRITIC, Eq.(20) is implemented as in Table 18.

$$w_l = \frac{\sqrt[3]{w_{l_{PSI}} \times w_{l_{ENT}} \times w_{l_{CRT}}}}{\sum_{j=1}^J \sqrt[3]{w_{j_{PSI}} \times w_{j_{ENT}} \times w_{j_{CRT}}}}, \sum_{m=1}^M w_l = 1.00 \quad (20)$$

Table 18. Aggregated weight for each criterion

OC_1	OC_2	OC_3	OC_4	OC_5
w_1	w_2	w_3	w_4	w_5
0.143	0.393	0.202	0.239	0.024

3.2. Computing criteria weights in SCG.

To compute criteria weights for SCG, RS, RR and RE methods were performed. 40 participants were ranked criteria in SCG from the most important to the least important. The first rank shows the most important criterion, the tenth rank presents the least important criterion. The part of criteria rankings belong to 40 participants are given in Table 18.

Table 18. Criteria rankings for 40 participants

P_j	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_{10}
	r_{1j}	r_{2j}	r_{3j}	r_{4j}	r_{5j}	r_{6j}	r_{7j}	r_{8j}	r_{9j}	r_{10j}
P_1	1	2	3	4	6	8	5	9	10	7
P_2	8	9	5	6	1	4	3	2	7	10
P_3	2	5	9	4	3	7	10	6	8	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
P_{38}	1	9	3	2	6	5	7	8	10	4
P_{39}	8	1	4	6	9	7	3	10	5	9
P_{40}	1	9	3	4	7	5	2	1	10	3

By using these rankings given in Table 18, RS was implemented to obtain subjective criteria weights ($w_{l_{RS}}$) as in Eq.(21).

$$w_{m_{jRS}} = \frac{M - r_{m_j} + 1}{\sum_{m=1}^M M - r_{m_j} + 1} \quad (21)$$

Here, $w_{m_{jRS}}$ is the weight of m th subjective criterion for j th participant according to RS. r_{m_j} is the rank of m th subjective criterion for j th participant. In Table 19, $w_{m_{jRS}}$ values are given.

Table 19. Each criterion weight for each participant according to RS

P_j	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_{10}
	$w_{1_{jRS}}$	$w_{2_{jRS}}$	$w_{3_{jRS}}$	$w_{4_{jRS}}$	$w_{5_{jRS}}$	$w_{6_{jRS}}$	$w_{7_{jRS}}$	$w_{8_{jRS}}$	$w_{9_{jRS}}$	$w_{10_{jRS}}$
P_1	0,041	0,050	0,038	0,028	0,020	0,014	0,027	0,010	0,005	0,020
P_2	0,012	0,011	0,028	0,020	0,040	0,032	0,037	0,043	0,018	0,005
P_3	0,037	0,033	0,009	0,028	0,032	0,018	0,005	0,024	0,014	0,050

⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
P_{38}	0,041	0,011	0,038	0,036	0,020	0,027	0,018	0,014	0,005	0,035
P_{39}	0,012	0,056	0,033	0,020	0,008	0,018	0,037	0,005	0,027	0,010
P_{40}	0,041	0,011	0,038	0,028	0,016	0,027	0,041	0,048	0,005	0,040

In the same manner, by using rankings in Table 18, RR was performed to compute criteria weights ($w_{m_{RR}}$) as in Eq.(22).

$$w_{m_{jRR}} = \frac{\frac{1}{r_{mj}}}{\sum_{m=1}^M \frac{1}{r_{mj}}} \quad (22)$$

Here, $w_{m_{jRR}}$ is the weight of m th criterion for j th participant according to RR. In Table 20, $w_{m_{jRR}}$ values are given.

Table 20. Each criterion weight for each participant according to RR

P_j	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_{10}
	$w_{1_{jRR}}$	$w_{2_{jRR}}$	$w_{3_{jRR}}$	$w_{4_{jRR}}$	$w_{5_{jRR}}$	$w_{6_{jRR}}$	$w_{7_{jRR}}$	$w_{8_{jRR}}$	$w_{9_{jRR}}$	$w_{10_{jRR}}$
P_1	0,067	0,052	0,036	0,028	0,012	0,012	0,018	0,009	0,008	0,014
P_2	0,008	0,011	0,022	0,012	0,074	0,024	0,030	0,040	0,012	0,009
P_3	0,033	0,021	0,012	0,017	0,025	0,014	0,009	0,013	0,011	0,095
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
P_{38}	0,067	0,011	0,036	0,035	0,012	0,019	0,013	0,010	0,008	0,024
P_{39}	0,008	0,103	0,027	0,012	0,008	0,014	0,030	0,008	0,017	0,011
P_{40}	0,067	0,011	0,036	0,017	0,011	0,019	0,046	0,080	0,008	0,032

Finally, RE was employed for criteria in SCG' rankings assigned by each participant as in Eq.(23).

$$w_{m_{jRE}} = \frac{(M - r_{mj} + 1)^z}{\sum_{m=1}^M (M - r_{mj} + 1)^z} \quad (23)$$

Here, $w_{m_{jRE}}$ is the weight of m th criterion for j th participant according to RE. In Table 21, $w_{m_{jRE}}$ values are given.

Table 21. Each criterion weight for each participant according to RE

P_j	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_{10}
	w_{1jRE}	w_{2jRE}	w_{3jRE}	w_{4jRE}	w_{5jRE}	w_{6jRE}	w_{7jRE}	w_{8jRE}	w_{9jRE}	w_{10jRE}
P_1	0,054	0,071	0,047	0,026	0,014	0,006	0,024	0,003	0,001	0,012
P_2	0,005	0,004	0,026	0,013	0,055	0,033	0,043	0,055	0,010	0,001
P_3	0,044	0,032	0,003	0,026	0,035	0,011	0,001	0,017	0,006	0,073
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
P_{38}	0,054	0,004	0,047	0,043	0,014	0,024	0,011	0,006	0,001	0,036
P_{39}	0,005	0,088	0,036	0,013	0,002	0,011	0,043	0,001	0,023	0,003
P_{40}	0,054	0,004	0,047	0,026	0,009	0,024	0,054	0,068	0,001	0,047

After computing criteria weights with three ranking-based weighting methods for each participant, aggregated weight for each criterion (w_m) given in Table 22 was obtained as in Eq.(24).

$$w_m = \frac{\sqrt[3]{w_{mjRS} \times w_{mjRE} \times w_{mjRR}}}{\sum_{j=1}^J \sqrt[3]{w_{mjRS} \times w_{mjRE} \times w_{mjRR}}}, \sum_{m=1}^M w_m = 1 \quad (24)$$

Table 22. Aggregated weight for each criterion

SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_{10}
w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
0.096	0.093	0.111	0.103	0.108	0.108	0.102	0.090	0.097	0.093

Fourth phase: Ranking websites by using CODAS considering OCG and SCG.

In this phase, different rankings for websites were determined in terms of objective and subjective usability criteria using CODAS method. Websites for OCG were evaluated on the basis of task and different rankings were obtained. An integrated evaluation was made by considering all criteria for SCG and a single ranking was obtained for the websites.

Step 4.1. Ranking websites for each task for OCG.

In this step, firstly, [X] and [N] matrices are created similar to the PSI method. Therefore, the matrices shown in Table 3 and Table 4 were used in CODAS calculations. Then, [N] is

weighted using Eq.(25) and weighted normalized initial decision matrix denoted as [R] is obtained as in Table 23.

$$r_{v_i} = w_l n_{v_i} \quad (25)$$

Here; r_{v_i} is the weighted normalized initial value of l th criterion in OCG for the v th website when performing i th task.

Then, the negative-ideal solution (point) is determined with Eq. (26) as in Table 23.

$$ns = [ns_{i_l}]_{1 \times l}$$

$$ns_{i_l} = \min_v r_{v_i} \quad (26)$$

Here; ns_{i_l} is the negative-ideal solution of l th criterion in OCG for i th task.

Euclidean and Taxicab distances of website alternatives from the negative-ideal solution are calculated for each task using Eq. (27) and Eq.(28). These values are shown Table 24.

$$E_{v_i} = \sqrt{\sum_{l=1}^L (r_{v_i} - ns_{i_l})^2} \quad (27)$$

$$T_{v_i} = \sum_{l=1}^L |r_{v_i} - ns_{i_l}| \quad (28)$$

Here; E_{v_i} and T_{v_i} are the Euclidean and Taxicab distances of v th website from the negative-ideal solution for i th task. To construct the relative assessment matrix (Ra) shown in Table 25, Eq. (29) and Eq.(30) are implemented.

$$Ra = [h_{vy}]_{V \times V} \quad (29)$$

$$h_{vy} = (E_{v_i} - E_{y_i}) + (\psi(E_{v_i} - E_{y_i})x(T_{v_i} - T_{y_i})) \quad (30)$$

where, $y \in \{1, 2, \dots, V\}$ and ψ is a threshold function to recognize the equality of the Euclidean distances of two websites as seen Eq.(31).

$$\psi(x) = \begin{cases} 1 & \text{if } |x| \geq \tau \\ 0 & \text{if } |x| < \tau \end{cases} \quad (31)$$

In this function, τ is a threshold parameter determined by decision maker. This parameter is recommended to take a value between 0.01 and 0.05. It is desired to have a value close to zero. If the difference between the Euclidean distances of the two website is greater than τ , these two websites are compared with the Taxicab distance.

Accordingly, the value of τ indicates a threshold function to explain the equality of Euclidean distances of the two alternatives. In this study, it was decided to determine the variance of Euclidean distances, unlike the traditional CODAS method, to define the equality of Euclidean distances. For this reason, it is proposed to calculate the variance (σ_i^2) value between Euclidean distances (E_{v_i}) by using Eq.(32) and to use this value instead of τ value.

$$\sigma_i^2 = \frac{\sum_{v=1}^v (E_{v_i} - \bar{E}_i)^2}{v - 1} \quad (32)$$

After obtaining σ_i^2 , the assessment score of each website (H_v) given in Table 25 are computed as in Eq.(33).

$$H_v = \sum_{y=1}^v h_{vy} \quad (33)$$

Table 23. Weighted normalized initial decision matrix and negative-ideal solution for OCG

WP _v	T ₁					T ₂					T ₃					T ₄					T ₅				
	OC ₁	OC ₂	OC ₃	OC ₄	OC ₅	OC ₁	OC ₂	OC ₃	OC ₄	OC ₅	OC ₁	OC ₂	OC ₃	OC ₄	OC ₅	OC ₁	OC ₂	OC ₃	OC ₄	OC ₅	OC ₁	OC ₂	OC ₃	OC ₄	OC ₅
	r _{v₁₁}	r _{v₁₂}	r _{v₁₃}	r _{v₁₄}	r _{v₁₅}	r _{v₂₁}	r _{v₂₂}	r _{v₂₃}	r _{v₂₄}	r _{v₂₅}	r _{v₃₁}	r _{v₃₂}	r _{v₃₃}	r _{v₃₄}	r _{v₃₅}	r _{v₄₁}	r _{v₄₂}	r _{v₄₃}	r _{v₄₄}	r _{v₄₅}	r _{v₅₁}	r _{v₅₂}	r _{v₅₃}	r _{v₅₄}	r _{v₅₅}
WP ₁	0,14	0,39	0,20	0,24	0,02	0,14	0,39	0,20	0,24	0,02	0,14	0,39	0,20	0,16	0,02	0,14	0,39	0,20	0,24	0,02	0,14	0,39	0,20	0,20	0,02
WP ₂	0,14	0,25	0,16	0,19	0,02	0,09	0,19	0,14	0,21	0,02	0,13	0,21	0,19	0,22	0,02	0,10	0,27	0,14	0,17	0,02	0,12	0,21	0,13	0,16	0,02
WP ₃	0,14	0,24	0,18	0,17	0,02	0,10	0,16	0,12	0,13	0,02	0,14	0,21	0,18	0,18	0,02	0,13	0,26	0,14	0,18	0,02	0,11	0,22	0,13	0,17	0,02
WP ₄	0,14	0,21	0,16	0,14	0,02	0,09	0,07	0,14	0,15	0,02	0,13	0,19	0,20	0,24	0,02	0,11	0,36	0,15	0,21	0,02	0,12	0,23	0,12	0,15	0,02
WP ₅	0,13	0,20	0,15	0,16	0,02	0,09	0,17	0,13	0,15	0,02	0,13	0,20	0,17	0,19	0,02	0,12	0,28	0,13	0,24	0,02	0,13	0,23	0,12	0,24	0,02
ns _i	0,13	0,20	0,15	0,14	0,02	0,09	0,07	0,12	0,13	0,02	0,13	0,19	0,17	0,16	0,02	0,10	0,26	0,13	0,17	0,02	0,11	0,21	0,12	0,15	0,02

Table 24. Euclidean and Taxicab distances for OCG

WP _v	T ₁		T ₂		T ₃		T ₄		T ₅	
	E _{v₁}	T _{v₁}	E _{v₂}	T _{v₂}	E _{v₃}	T _{v₃}	E _{v₄}	T _{v₄}	E _{v₅}	T _{v₅}
WP ₁	2E-02	3E-01	6E-02	6E-01	2E-02	2E-01	1E-02	3E-01	2E-02	3E-01
WP ₂	3E-03	1E-01	1E-02	2E-01	2E-03	1E-01	5E-05	1E-02	1E-04	3E-02
WP ₃	2E-03	1E-01	4E-03	1E-01	4E-04	5E-02	4E-04	4E-02	1E-04	2E-02
WP ₄	2E-04	3E-02	5E-04	4E-02	3E-03	1E-01	5E-03	2E-01	2E-04	2E-02
WP ₅	4E-04	3E-02	6E-03	1E-01	4E-04	4E-02	2E-03	9E-02	4E-03	1E-01

Table 25. The relative assessment matrix and assessment scores of web pages respect to OCG

WP _v	T ₁								T ₂								T ₃							
	WP ₁	WP ₂	WP ₃	WP ₄	WP ₅	H _v	R	WP ₁	WP ₂	WP ₃	WP ₄	WP ₅	H _v	R	WP ₁	WP ₂	WP ₃	WP ₄	WP ₅	H _v	R			
WP ₁	0,0E+00	2,5E-01	2,7E-01	3,4E-01	3,4E-01	1,2E+00	1	1,3E-01	4,1E-01	5,3E-01	5,9E-01	4,9E-01	2,2E+00	1	0,0E+00	1,6E-01	2,2E-01	1,7E-01	1,7E-01	7,3E-01	1			
WP ₂	-2,5E-01	0,0E+00	1,5E-02	8,5E-02	8,9E-02	-6,4E-02	5	-2,8E-01	0,0E+00	1,2E-01	1,8E-01	8,5E-02	1,0E-01	2	-1,6E-01	0,0E+00	5,6E-02	6,5E-03	9,2E-03	-9,2E-02	5			
WP ₃	-2,7E-01	-1,5E-02	0,0E+00	7,1E-02	7,4E-02	-1,4E-01	2	-4,0E-01	-1,2E-01	0,0E+00	6,3E-02	-3,4E-02	-4,9E-01	4	-2,2E-01	-5,6E-02	0,0E+00	-5,0E-02	-1,3E-05	-3,3E-01	4			
WP ₄	-3,4E-01	-8,5E-02	-7,1E-02	0,0E+00	3,7E-03	-4,9E-01	3	-4,6E-01	-1,8E-01	-6,3E-02	0,0E+00	-9,7E-02	-8,0E-01	3	-1,7E-01	-6,5E-03	5,0E-02	0,0E+00	2,7E-03	-1,2E-01	2			
WP ₅	-3,4E-01	-8,9E-02	-7,4E-02	-3,7E-03	0,0E+00	-5,1E-01	4	-3,7E-01	-8,5E-02	3,4E-02	9,7E-02	0,0E+00	-3,2E-01	5	-2,1E-02	-6,9E-02	1,3E-05	-6,2E-02	0,0E+00	-1,5E-01	3			
WP _v	T ₄								T ₅															
	WP ₁	WP ₂	WP ₃	WP ₄	WP ₅	H _v	R	WP ₁	WP ₂	WP ₃	WP ₄	WP ₅	H _v	R										
WP ₁	0,0E+00	3,1E-01	2,8E-01	1,6E-01	1,2E-02	7,5E-01	1	0,0E+00	3,4E-01	3,4E-01	3,4E-01	2,3E-01	1,2E+00	2										
WP ₂	-3,1E-01	0,0E+00	-2,9E-02	-1,5E-01	-2,1E-03	-4,9E-01	5	-3,4E-01	0,0E+00	-2,5E-06	-3,8E-05	-1,0E-01	-4,4E-01	3										
WP ₃	-2,8E-01	2,9E-02	0,0E+00	-1,2E-01	-1,7E-03	-3,7E-01	4	-3,4E-01	2,5E-06	0,0E+00	-3,5E-05	-1,1E-01	-4,5E-01	4										
WP ₄	-1,6E-01	1,5E-01	1,2E-01	0,0E+00	3,1E-03	1,2E-01	2	-3,4E-01	3,8E-05	3,5E-05	0,0E+00	-1,1E-01	-4,5E-01	5										
WP ₅	-2,3E-01	7,7E-02	4,8E-02	-7,3E-02	0,0E+00	-1,8E-01	3	-2,3E-01	1,0E-01	1,1E-01	1,1E-01	0,0E+00	9,0E-02	1										

Step 4.2. Ranking websites for SCG.

In this step, firstly, initial decision matrix [X] is structured for SCG as seen Table 26. [X] includes rank values of each website according to SUS. Each element of [X] is denoted as x_{vm} which values were obtained by mode of $x_{j_{vm}}$ values as seen Table 26. $x_{j_{vm}}$ is the rank of m th criterion for j th participant on the v th website.

Table 26. Initial decision matrix for SCG

	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_{10}
	x_{v_1}	x_{v_2}	x_{v_3}	x_{v_4}	x_{v_5}	x_{v_6}	x_{v_7}	x_{v_8}	x_{v_9}	$x_{v_{10}}$
WP_1	4	1	4	4	5	3	5	1	3	2
WP_2	5	1	4	1	3	1	4	1	4	1
WP_3	5	3	5	1	3	3	5	1	3	1
WP_4	2	1	5	1	3	3	3	1	4	3
WP_5	5	3	3	3	4	1	3	1	5	1

Then, [N] is obtained using Eq. (34) and Eq. (35) as shown Table 27. Among criteria in SCG, the SC_2, SC_4, SC_6, SC_8 and SC_{10} criteria are cost type.

For benefit type criteria in SCG,

$$n_{vm} = \frac{x_{vm}}{x_{m_{max}}} \quad (34)$$

For cost type criteria in SCG,

$$n_{vm} = \frac{x_{m_{min}}}{x_{vm}} \quad (35)$$

Table 27. Normalized initial decision matrix for SCG

	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_{10}
	n_{v_1}	n_{v_2}	n_{v_3}	n_{v_4}	n_{v_5}	n_{v_6}	n_{v_7}	n_{v_8}	n_{v_9}	$n_{v_{10}}$
WP_1	0,80	1,00	0,80	0,25	1,00	0,33	1,00	1,00	0,60	0,50
WP_2	1,00	1,00	0,80	1,00	0,60	1,00	0,80	1,00	0,80	1,00
WP_3	1,00	0,33	1,00	1,00	0,60	0,33	1,00	1,00	0,60	1,00
WP_4	0,40	1,00	1,00	1,00	0,60	0,33	0,60	1,00	0,80	0,33
WP_5	1,00	0,33	0,60	0,33	0,80	1,00	0,60	1,00	1,00	1,00

Normalized initial decision matrix is weighted using Eq.(36) and [R] are given in Table 28.

$$r_{v_m} = w_m n_{v_m} \quad (36)$$

Table 28. Weighted normalized initial decision matrix for SCG

	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_{10}
	r_{v_1}	r_{v_2}	r_{v_3}	r_{v_4}	r_{v_5}	r_{v_6}	r_{v_7}	r_{v_8}	r_{v_9}	$r_{v_{10}}$
WP_1	0,08	0,09	0,09	0,03	0,11	0,04	0,10	0,09	0,06	0,05
WP_2	0,10	0,09	0,09	0,10	0,06	0,11	0,08	0,09	0,08	0,09
WP_3	0,10	0,03	0,11	0,10	0,06	0,04	0,10	0,09	0,06	0,09
WP_4	0,04	0,09	0,11	0,10	0,06	0,04	0,06	0,09	0,08	0,03
WP_5	0,10	0,03	0,07	0,03	0,09	0,11	0,06	0,09	0,10	0,09

Additionally, the negative-ideal solution (point) is determined with Eq.(37) and Eq.(38) as seen Table 28.

$$ns = [ns_m]_{1 \times m} \quad (37)$$

$$ns_m = \min_v r_{v_m} \quad (38)$$

According to computed r_{v_m} and ns_m values, the Euclidean (E_v) and Taxicab (T_v) distances of alternatives from the negative-ideal solution are calculated for each task, shown as Eq.(39) and Eq.(40). These values are seen in Table 29.

$$E_v = \sqrt{\sum_{m=1}^M (r_{v_m} - ns_m)^2} \quad (39)$$

$$T_v = \sum_{m=1}^M |r_{v_m} - ns_m| \quad (40)$$

Table 29. Euclidean and Taxicab distances for SCG

	E_v	T_v
WP_1	0,00	0,22
WP_2	0,01	0,39
WP_3	0,01	0,28
WP_4	0,01	0,20
WP_5	0,01	0,26

To construct the relative assessment matrix shown in Table 30, Eq.(29) and Eq.(41) are implemented.

$$h_{vy} = (E_v - E_y) + (\psi(E_v - E_y)x(T_v - T_y)) \quad (41)$$

In Eq.(41), variance which is calculated Eq. (42) is used instead of τ value.

$$\sigma^2 = \frac{\sum_{v=1}^V (E_v - \bar{E}_m)^2}{v - 1} \quad (42)$$

After obtaining σ^2 , the assessment score of each alternative given in Table 30 are computed as in Eq.(42).

Table 30. The relative assessment matrix and assessment scores of websites respect to SCG

	WP_1	WP_2	WP_3	WP_4	WP_5	H_v	R
WP_1	0,00	-0,18	-0,06	0,02	-0,04	-0,26	4
WP_2	0,18	0,00	0,11	0,20	0,14	0,62	1
WP_3	0,06	-0,11	0,00	0,08	0,02	0,05	5
WP_4	-0,02	-0,20	-0,08	0,00	-0,06	-0,35	3
WP_5	0,04	-0,14	-0,02	0,06	0,00	-0,06	3

4. RESULTS

In this study, to overcome shortcomings related to traditional usability evaluation techniques mentioned in introduction section, a new hybrid MDUEA was suggested to compare usability levels of AFWSs. The advanced approach includes both user tests data obtained from Morae V3, SUS results and users' ranking results related to SCG. PIS, ENTROPY and CRITIC methods were performed to find criteria weights for OCG. RS, RR, RE methods were utilized to find criteria weights for SCG. The methods for determining objective and subjective criteria weights were chosen according to these criteria' natures. PSI, ENTROPY and CRITIC weighting methods work with quantitative data and OCG considered in this study also have quantitative data computed by Morae V3. RS, RR, RE methods are suitable for criteria having subjective nature. RS, RR, RE are ranking based weighting methods and these methods consider criteria rankings identified by end users then, these rankings are converted to weights. In this manner, both quantitative values of criteria and end users' opinions were reflected to criteria weights. According to PSI results, success percentage was determined as the most important criterion, task time was identified as the least important criterion. Results obtained from ENTROPY showed that time between two data entries is the most important while success percentage is the least important for end users. In term od CRITIC's results, time between two

data entries was found as the most important and success percentage was evaluated as the least important criterion for end users. According to the aggregated results, time between two data entries has the highest importance weight while success percentage has the lowest weight. When these results are evaluated it was seen that ENTROPY, CRITIC and aggregated approaches give the same results. It is considered that this situation stems from the focus on the knowledge level related to criteria in ENTROPY and CRITIC methods. Accordingly, two of the three criteria weighting methods reflected more than to the aggregated weight results, because they did the weighting using similar strategies. In term of SCG, easy to use level was determined as the most important criterion and level of inconsistency was found as the least important one for end users. During weighting procedure od SCG, each participant's ranking from the most important to least important was considered and RS, RR and RE approaches were utilized and aggregated weights were computed for each participant. Then aggregated geometric weight was calculated for all participant as a final weight of each criterion in SCG. This type of weight evaluation was performed at first in this study.

For managerial implication obtained from the study, it can be said that the produce results from the proposed approach are logical. For the OCG, time between two data entries was determined as the most important criterion for end users. The time between two data entries also indicates the time the end user spends moving on to the next step, indicating how well the user knows the website, as well as whether the website is directing the user correctly. If the website informs the user about the next step correctly, the user will reach his goal in a short time by making the other move. Otherwise, the user will have to think about his next step. This criterion is important in determining how soon the user will reach his aim. The fact that this criterion is the most important brought about a decrease in the importance of the success percentage. Because, when the time between two data entries is shortened, the percentage of success increases. In this situation, user can quickly reach his goal and complete the task correctly. In this case, the user considers the usability of the website to be high. This evaluation is obtained from SUS. In term of SCG, it is logical to find easy to use level as the most important criterion for users. An easy-to-use website can get the user to the goal as soon as possible. The user can perform all tasks without difficulty. As with the same objective criteria, the high importance of easy use criterion in SCG has decreased the importance of the level of inconsistency criterion. If the user can use the website easily, it is directed by the website with a logical order of the steps to do in every task. This will result in a decrease in the inconsistency level. When the weights of objective and subjective criteria are examined, it is seen that the most important criteria in both groups are in

a logical harmony. The short time between two data entries means that the website can be used easily. The user knows the next step or is directed by the website correctly.

Additionally, a new version of CODAS was suggested in the study. Determining threshold function value in an objective manner was first developed in this research. In this way, subjectivity in the implementation of the CODAS was overcome. However, website alternatives are evaluated separately using the CODAS method considering OCG and SCG. CODAS is used the Euclidean distance as the primary and the Taxicab distance as the secondary measure to determine the desirability of an alternative. These distances are calculated according to the negative-ideal point and the alternative which has greater distances is more desirable in the CODAS method. In this study, the alternative farthest from the negative ideal point has been accepted as the best alternative in terms of usability. Because it is the alternative that does not satisfy the alternative criteria close to the negative ideal solution at the desired level. As the distance to this point increases, the satisfaction level of the criteria increases and the alternative that provides this satisfaction level becomes the best alternative.

Three-dimension initial decision matrices for OCG and SCG were structured and CODAS was performed for these two separate three-dimension initial decision matrices firstly. Briefly, AFWSs were ranked for OCG and SCG separately. In this way, AFWSs can be evaluated in terms of OCG and SCG separately. This provides managers to see the location of their firms' websites for these two different criteria groups. In OCG, WP_1 is determined as the best website for the other tasks excluding T_5 . WP_1 is found more usable than other airline website for doing selecting departure and arrival flights and purchasing tickets, online check in, seat selection and baggage tracking tasks. WP_5 seems to be more useful in learning flight status task than other websites. In SCG, WP_2 is determined as the best website WP_2 is found more usable and satisfying than other AFWSs while doing tasks.

5. COCNLUSIONS, DISCUSSION AND FUTURE STUDIES

In recent years, airlines tend to use the internet and information technologies more to gain competitive advantage against other companies. In particular, these companies see their websites as an important tool to advertise, serve customers directly without intermediaries or travel agencies, reduce costs and attract new customers. Customers can perform many operations such as selecting departure and arrival flights and purchasing tickets, online check-in, seat selection, baggage tracking and learning flight status using the websites of airline

companies. Websites must be available for these operations to be carried out easily, quickly and smoothly. An unavailable website may cause customers to change their company preferences. In this study, a hybrid MDUEA was proposed to evaluate airline websites in terms of usability. Unlike the MCDM approaches recommended for usability assessment in the literature, in this approach, the results of usability testing with Morae V3 package program and the results of SUS were combined by using MCDM structure. Also, a three-dimension initial decision matrix was advanced in MCDM approach includes tasks, websites and usability criteria at the same time. Subjective and objective criteria were considered to evaluate usability of AFWSs. Usability criteria weights were computed by implementing different weighting approaches which are suitable for subjective and objective criteria. Additionally, the weight results of these approaches were aggregated. For the OCG, websites were evaluated on the basis of tasks and different rankings were obtained. For the SCG, an integrated evaluation was made by considering all criteria and a single ranking was obtained for the websites.

For future studies, experts as well as users can evaluate websites in terms of usability. In this context, by combining usability test and heuristic evaluation results, criteria weights and website priorities can be determined more accurately and effectively. In addition, universal design principles can be evaluated together with usability criteria. In addition, the effects of users' mental workloads related to usability of websites can be taken into account. The proposed approach can be expanded using different MCDM techniques with hesitant fuzzy numbers to consider the instability caused by human nature and the uncertainty of the decision environment.

Ethical approval

There is no need for ethical approval for this study. All context of the study is belong to the authors.

Funding details (In case of Funding)

There is no funding related to this study.

Conflict of interest

We have not any conflict of interests.

Informed Consent

There is no need for informed consent related to this study. All context of the study is belong to the authors.

Author's contribution

Assc.Prof.Dr.Gülin Feryal CAN determined research gap and she suggested the novel multi-dimensional usability evaluation approach. She wrote criteria weighting part of the proposed approach's algorithm.

Assc.Prof.Dr.Elif KILIÇ DELİCE searched for literature related to the airline websites' usability evaluations and MCDM based usability evaluations. She wrote websites ranking phase of the algorithm.

Introduction, conclusions, discussions and future research directions were written by both of the authors.

REFERENCES

Akincilar A, Dagdeviren M. (2014) A hybrid multi-criteria decision making model to evaluate hotel websites. *International Journal of Hospitality Management* 36: 263-271.

Attri, R, Grover S. (2015) Application of preference selection index method for decision making over the design stage of production system life cycle. *Journal of King Saud University-Engineering Sciences* 27(2):207-216.

Brooke J. (1996) SUS - A quick and dirty usability scale. *usability evaluation in industry*, 189(194): 4-7. <https://doi.org/10.1002/Hbm.20701> Accessed 26 June 2020

Cao K, Yang Z. (2016) A study of e-commerce adoption by tourism websites in China. *Journal of Destination Marketing & Management* 5(3):283-289.

Cebi S. (2013) Determining importance degrees of website design parameters based on interactions and types of websites. *Decision Support Systems*, 54(2):1030-1043.

Chen M. (2018) Improving website structure through reducing information overload. *Decision Support Systems*, 110:84-94.

Chmielarz W, Zborowski M. (2018) Analysis of e-banking websites' quality with the application of the TOPSIS method—a practical study. *Procedia computer science*, 126:1964-1976.

Díaz E, Martín-Consuegra D. (2016) A latent class segmentation analysis of airlines based on website evaluation. *Journal of Air Transport Management*, 55: 20-40.

- Díaz J, Rusu C, Collazos C. A. (2017) Experimental validation of a set of cultural-oriented usability heuristics: e-Commerce websites evaluation. *Computer Standards & Interfaces*, 50: 160-178.
- Diakoulaki D, Mavrotas G, Papayannakis L. (1995) Determining objective weights in multiple criteria problems: The CRITIC method. *Computers & Operations Research*, 22(7): 763-770.
- Gutierrez C. F, Loucopoulos C, Reinsch R. W. (2005) Disability-accessibility of airlines' Web sites for US reservations online. *Journal of Air Transport Management*, 11(4): 239-247.
- Harison E, Boonstra A. (2008) Reaching new altitudes in e-commerce: assessing the performance of airline websites. *Journal of Air Transport Management*, 14(2):92-98.
- Ilbahar E, Cebi S. (2017) Classification of design parameters for E-commerce websites: A novel fuzzy Kano approach. *Telematics and Informatics*, 34(8):1814-1825.
- Kang D, Jang W, Park Y. (2016) Evaluation of e-commerce websites using fuzzy hierarchical TOPSIS based on ES-QUAL. *Applied Soft Computing*, 42:53-65.
- Karkin N, Janssen M. (2014) Evaluating websites from a public value perspective: A review of Turkish local government websites. *International Journal of Information Management*, 34(3): 351-363.
- Ghorabae M. K, Zavadskas E. K, Turskis, Z, Antucheviciene, J. (2016) A new combinative distance-based assessment (CODAS) method for multi-criteria decision-making. *Economic Computation & Economic Cybernetics Studies & Research*, 50(3):25-44.
- Lee-Geiller S, Lee T. D. (2019) Using government websites to enhance democratic E-governance: A conceptual model for evaluation. *Government Information Quarterly*, 36(2): 208-225.
- Liang D, Zhang Y, Xu Z, Jamaldeen A. (2019). Pythagorean fuzzy VIKOR approaches based on TODIM for evaluating internet banking website quality of Ghanaian banking industry. *Applied Soft Computing*, 78: 583-594.
- Lituchy T. R, Barra R. A. (2008) International issues of the design and usage of websites for e-commerce: Hotel and airline examples. *Journal of Engineering and Technology Management*, 25(1-2): 93-111.

Storto C.L. (2013) Evaluating ecommerce websites cognitive efficiency: An integrative framework based on data envelopment analysis. *Applied ergonomics*, 44(6): 1004-1014.

Malczewski J. (1999) *GIS and multicriteria decision analysis*. John Wiley & Sons.

Maniya K, Bhatt M. G. (2010) A selection of material using a novel type decision-making method: Preference selection index method. *Materials & Design*, 31(4): 1785-1789.

Nielsen J. (1993). *Usability Engineering*. Morgan Kaufmann.

Ostovare M, Shahraki M. R. (2019) Evaluation of hotel websites using the multicriteria analysis of PROMETHEE and GAIA: Evidence from the five-star hotels of Mashhad. *Tourism Management Perspectives*, 30:107-116.

Pamučar D, Stević Ž, Zavadskas E. K. (2018) Integration of interval rough AHP and interval rough MABAC methods for evaluating university web pages. *Applied Soft Computing*, 67: 141-163.

TechSmith. (2007) Technical Overview of Morae: Usability testing software with insight you can share. <https://docplayer.net/18811235-Technical-overview-of-morae-usability-testing-software-with-insight-you-can-share.html> / Accessed 13 October 2020.

Tsai W H, Chou W C, Lai C W. (2010) An effective evaluation model and improvement analysis for national park websites: A case study of Taiwan. *Tourism Management*, 31(6):936-952.

Tsai W H, Chou W C, Leu J. D. (2011) An effectiveness evaluation model for the web-based marketing of the airline industry. *Expert Systems with Applications*, 38(12):15499-15516.

Verkijika S F, De Wet L. (2018) A usability assessment of e-government websites in Sub-Saharan Africa. *International Journal of Information Management*, 39: 20-29.

Wu J, Sun, J, Liang L, Zha Y. (2011) Determination of weights for ultimate cross efficiency using Shannon entropy. *Expert Systems with Applications*, 38(5): 5162-5165.

Appendix: SUS

Strongly
disagree

Strongly
agree

1. I think that I would like to use this system frequently

1	2	3	4	5

2. I found the system unnecessarily complex

1	2	3	4	5

3. I thought the system was easy to use

1	2	3	4	5

4. I think that I would need the support of a technical person to be able to use this system

1	2	3	4	5

5. I found the various functions in this system were well integrated

1	2	3	4	5

6. I thought there was too much inconsistency in this system

1	2	3	4	5

7. I would imagine that most people would learn to use this system very quickly

1	2	3	4	5

8. I found the system very cumbersome to use

1	2	3	4	5

9. I felt very confident using the system

1	2	3	4	5

10. I needed to learn a lot of things before I could get going with this system

1	2	3	4	5

Figures

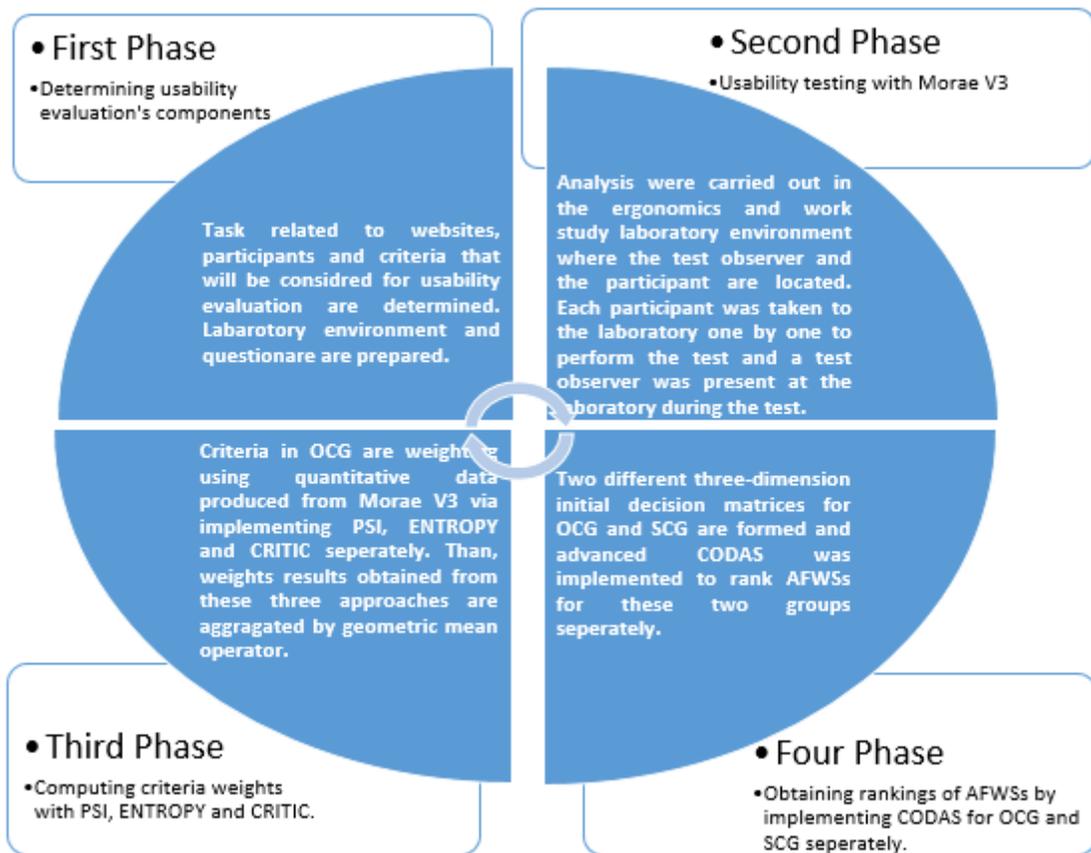


Figure 1

The flow chart of the proposed approach