

# A machine learning-based usability evaluation method for eLearning systems

Asil Oztekin<sup>a</sup>, Dursun Delen<sup>b,\*</sup>, Ali Turkyilmaz<sup>c</sup>, Selim Zaim<sup>d</sup>

<sup>a</sup> Department of Operations and Information Systems, Manning School of Business, University of Massachusetts Lowell, Lowell, MA 01854, USA

<sup>b</sup> Department of Management Science and Information Systems, Spears School of Business, Oklahoma State University, Tulsa, OK 74106, USA

<sup>c</sup> Department of Industrial Engineering, College of Engineering, Fatih University, Buyukcekmece, 34500 Istanbul, Turkey

<sup>d</sup> Department of Mechanical Engineering, College of Technology, Marmara University, Kadikoy, 34722 Istanbul, Turkey

## ARTICLE INFO

### Article history:

Received 21 May 2012

Received in revised form 22 April 2013

Accepted 9 May 2013

Available online 17 May 2013

### Keywords:

eLearning (web-based learning/distance learning)

Usability engineering

Severity index

Information fusion

Sensitivity analysis

Machine learning

## ABSTRACT

The research presented in this paper proposes a new machine learning-based evaluation method for assessing the usability of eLearning systems. Three machine learning methods (support vector machines, neural networks and decision trees) along with multiple linear regression are used to develop prediction models in order to discover the underlying relationship between the overall eLearning system usability and its predictor factors. A subsequent sensitivity analysis is conducted to determine the rank-order importance of the predictors. Using both sensitivity values along with the usability scores, a metric (called severity index) is devised. By applying a Pareto-like analysis, the severity index values are ranked and the most important usability characteristics are identified. The case study results show that the proposed methodology enhances the determination of eLearning system problems by identifying the most pertinent usability factors. The proposed method could provide an invaluable guidance to the usability experts as to what measures should be improved in order to maximize the system usability for a targeted group of end-users of an eLearning system.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

eLearning (a.k.a. Open and Distance Learning [ODL] or web-based learning) means *electronic learning* that utilizes electronic communication for teaching and learning designed to be applied from a distance. eLearning can be as effective as the conventional in-class face-to-face teaching and learning, if the techniques are appropriate for the teaching goals with a well-organized student–teacher interaction. Both users and creators of eLearning systems would request effective and efficient teaching and learning products and services, all of which are referred to as “quality” [24].

As in other web-based systems, eLearning system end-users also request more *usable* and high *quality* systems. Methods to measure the quality of computer systems have been considered first under the heading of *ergonomics* and *ease-of-use*, and later under the heading of *usability* [25]. Usability is basically defined as “ease of use” which can apply to practically any object that is used for some purpose [46]. It has also been defined as the extent to which an application allows users to accomplish specified goals efficiently, effectively, and satisfactorily [47]. Formally defined, *usability* stands for “the effectiveness, efficiency, and satisfaction with which specified users can

achieve goals in particular environments” [25]; “the capability to be used by humans easily and effectively”; “quality in use” [6,7]; “how easy it is to find, understand and use the information displayed on a web-based system” [35]; and “the ultimate quality factor” for software architecture [70].

The International Standardization Organization (ISO) defines usability as “the extent to which a product can be used by specified users to achieve specified goals with efficiency, effectiveness, and satisfaction in a specified context of use” [30]. Especially in software development, usability evaluation is listed among the foremost strategic factors to be confronted and therefore has been an increasingly popular topic of the human–computer interaction area [31,33]. There has been a perception that usability and quality are somehow related to each other [29,31,32,52]. Which one affects the other and in what ways are the debates that may vary from one research domain to another one. To exemplify, Oztekin et al. [56] conceptually proved that *quality* is a subset of *usability* in the case of usability evaluation of web-based information systems. The conventional approaches have conducted usability evaluation through design heuristics [51] and usability rules [69]. Although these techniques deserve merit and base the fundamentals of the usability evaluation, they lack a quantitative basis for usability evaluation. However, the fact that the eLearning system end-users request more usable and quality systems necessitates a sophisticated quantitative methodology to handle the usability-related problems so as to satisfy eLearning end-users.

\* Corresponding author. Tel.: +1 918 594 8283.

E-mail address: [dursun.delen@okstate.edu](mailto:dursun.delen@okstate.edu) (D. Delen).

## 2. Literature review

In the current state-of-the-art, several attempts exist that evaluate the usability of various systems [3,10,12,21,37,42,44,45,64]. The methodologies proposed in these studies have been widely accepted and applied in a wide variety of fields. However, they have some weaknesses and hence require improvements in some aspects. For example, they either do not present a quantitative methodology to be adopted along with their usability checklists, or do not compute a comprehensive objective score that represents the overall usability index. They mainly focus on addressing the limited number of usability problems in a qualitative manner by ignoring the internal effect of many other usability-related factors. However, in reality, usability-related improvements often require a cyclic process of reassessment–retesting–improvement of the underlying eLearning system.

Using the expectancy disconfirmation theory (EDT), Chiu et al. [13] model the users' intention to continue using an eLearning system. This study is the leading endeavor to model the eLearning usability in a cause-and-effect manner. Nevertheless, it does not explain the reasons why the perceived usability or quality is not satisfactory enough. Therefore, such an elaboration still remains critical in order to improve the system so as to satisfy the end-users. Downey et al. [20] search the effect of national culture on usability perception regarding the eLearning. Forming representative pools of national cultures, the most striking conclusion they reached was that cultures with high power distance indicators (e.g. China, India, Indonesia, Malaysia, and Singapore) tend to have more collectivist rather than individualistic tendencies. The individuals from these cultures had a strong satisfaction level with the eLearning system that they used. In order to understand the potential of technology-mediated learning (or online learning) for supporting students' learning, Hu and Hui [28] examine how it influences the underlying learning process as well as the students' learning effectiveness and satisfaction, compared to the face-to-face learning. The results from an experiment involving 212 university students indicate the benefits and constraints of technology-mediated learning, together with the need for proper design and selection of technology-mediated learning systems and teaching strategies to improve students' learning effectiveness and satisfaction. Hsu et al. [27] develop some design criteria and a usability evaluation scale for eLearning systems. They suggest four dimensions (i.e. instructional strategy, teaching material, learning tool, and learning interface) for the usability evaluation scale and by using these they compare three different eLearning websites. This study presents a comprehensive usability evaluation scale, but it does not provide an analytical basis to enlighten why one eLearning website is superior to another, which is actually substantial information to improve other eLearning systems further.

Parlangeli et al. [58] assessed the effect of usability on the learning of the students through a three-step evaluation method. The first step was conducted by a heuristic evaluation of two experts in the human-computer interaction area. The second step was a user-testing performed by ten college students, and the third step is performed by thirty-six high school students. The study concluded that the hypertexts can make the user feel lost and the problem is more severe if the user is not familiar with the topic. This study even itself admits that more quantitative results should be presented to validate the main hypothesis underlying the study. Reviewing the existing usability checklists, Squires and Preece [71] criticize them in terms of the lack of learning perspective. To fill this gap, they propose a heuristics evaluation-based usability predictive methodology which is an extension of Molich and Nielsen's [49]. Although this study presents a broader perspective for the eLearning software usability testing, no case study is provided for a quantitative validation. Arbaugh and Benbunan-Fich [1] consider participant interaction as one of the strongest predictors of success in the online learning environments. Along with their research findings, the authors show that students in collaborative courses experience

higher levels of learner–learner and learner–system interaction, while only learner–instructor and learner–system interaction are significantly associated with higher perceptions of learning. Sun et al. [73] indicate that the most critical functionalities of a successful eLearning system are instruction presentation, including e-syllabus and electronic whiteboard, and student learning management, which can be achieved by online discussion, online roll call, and assignment management. By applying a series of surveys on fifty-four college students, Storey et al. [72] identify the expectations from an eLearning system. This study illustrates a framework for students and instructors, who use eLearning systems, yet it also lacks an analytical model which would explain why the usability is good or bad and what criteria affect it, to what degree. A more recent study conducted by Ben Ayed et al. [5] develops a decision support system based on knowledge discovery from databases to propose a user-centered approach in medical field. Although this study incorporates DSS approach, it lacks quantitative modeling perspective in essence. Earlier, Web-based DSS was also discussed in the study by Bharati and Chaudhury [8], where they investigated the factors such as information quality and system quality that impact decision-making satisfaction in web-based decision support systems.

Based on the abovementioned, the state-of-the-art reveals that there is a gap in the literature to determine the main usability problems and provide a rank order to be dealt with sequentially. If an eLearning system has some usability problems, the cause of this low usability should be identified and improved. These problems call for the implementation of an analytical approach where relevant metrics are used to identify the most important usability-related factors efficiently and effectively, so that the limited resources such as time and money available for improving the eLearning system are optimally utilized. This study proposes such a systematic methodology. The details of the proposed methodology are further explained in Section 3.

## 3. Proposed methodology

There are a number of usability questionnaires/checklists proposed in the literature, including Sumi [36,37], Quis [53], PSSUQ [40], and PutQ [41]. A more recent one has been proposed by Lee and Kozar [39] which examines the causal relationship between the website usability and online purchase intent. Thereafter, they provide insights for improving the website usability in order to positively affect the online customer behavior. These instruments derive their content and structure from the leading assessment studies mentioned in the previous section.

The fundamental limitation of the above mentioned instruments is the fact that they merely employ a checklist-type usability testing, developed after the opinions of some usability experts or test participants. They do not provide an analytical foundation and numerical evidence which would rank the emerging usability items in terms of their importance for further improvement and remedy. They merely rely on the evaluation results provided by a representative sample pool of intended end-users or domain experts and rank their evaluation results based on the mean scores received for each checklist item. What they mainly overlook is that whether or not improving a particular usability problem would eventually have a significant effect on the usability perception of the end-users. In other words, they reveal the problematic usability aspects of the eLearning system by suggesting that the smaller the survey-based evaluation average score, the more important the checklist item is; hence the priority should be given to it for improvement. Yet, they ignore the effect of one unit change of this particular checklist item on the final usability perception. Although one usability checklist item is fairly low (evaluated badly by the end-users or usability experts), it may or may not make a big improvement on the final usability of the eLearning system analyzed. Since time and efforts should be focused only on worthy usability problems that

would make a significant difference on the usability at the end, both values should be considered in the usability evaluation process. Therefore, in this study we define a combined metric, *severity index*, which takes both of the measures into account and can be calculated as in Eq. (1).

$$\text{Severity index} = \text{Sensitivity score} * \frac{1}{\text{Average of checklist evaluation scores}} \quad (1)$$

Where, sensitivity score refers to the importance of the checklist item in predicting the usability and will be explained in Section 3.2.3 in more detail. On the other hand, average of checklist evaluation scores is straightforward but the reason to take the reciprocal of it is as follows. The smaller the average of checklist evaluation scores for an item, the more important that checklist item is. However, the bigger the sensitivity score, the more important the checklist item is. To make them comparable and combined in one metric, namely in *sensitivity index*, we propose to use the inverse of the former. The overall methodology is illustrated as in Fig. 1.

The first step in the proposed methodology is to collect sample data from representative end-users to be able to apply *usability testing* process. The data is collected using a survey instrument, UseLearn checklist [55], that was developed based on the checklists and questionnaires in the published literature which was mentioned in Section 1. Table 1 summarizes the UseLearn checklist items and their corresponding questions. For further usage throughout the paper, we refer the checklist questions/items as the input variables to predict the overall usability of the eLearning system, and we prefer doing this by means of their abbreviated symbols such as EP1, EP2, and EP3 for the first, second, and third questions related to *error prevention* factor of the UseLearn checklist.

The output variable (overall usability of the eLearning system) is represented as the last question in this checklist. We used a 5-point Likert scale, with anchors ranging from strongly disagree (1) to strongly agree (5). Therefore, as a threshold for “satisfactory usability score”, we adopted the Likert point of 4 (equivalent to *agree*). This is translated

into that the checklist question measuring the output variable (the overall eLearning system usability) should be rated 4 or greater on average to be able to call the eLearning system usability “satisfactory”. If so, the end-users are assumed to find the system usable enough; thus no further analysis is required. Otherwise, the evaluation process proceeds as follows. Since the complex relationship between the output variable and the input variables are not known a priori, the next step is to find the best predictive model that explains this potentially complex relationship considering various performance measures (i.e. error rates and correlation). Using this predictive model, a sensitivity analysis is conducted for the input variables (checklist items). Then, the *severity index* is calculated for the checklist items to rank them in descending order. This ranking would help determine which variables to improve first with scarce time and money constraints in hand. A *pseudo-Pareto chart* would be useful to pick only the most effective checklist items and corresponding usability problems of the eLearning system. The Pareto rule [57] basically claims that 80% of the problems stem from 20% of the causes. Therefore, instead of dealing with all root causes on hand; it suggests handling 20% of them which would hypothetically help solve 80% of the problems. In this way, the severest usability problems can be revealed and improved in order to improve the overall usability. As seen in Fig. 1, usability evaluation algorithm is actually a loop which would be terminated when the desired level of usability level is achieved.

### 3.1. Prediction models

The relationships between the output variable and the input variables are to be explained by the prediction models. We used four different models to reveal these relations: multiple linear regression, decision trees, neural networks, and support vector machines. This set of predictive models was selected due to their popularity in the literature. The general purpose of *multiple linear regression* [59] is to learn more about the relationship between several independent variables (a.k.a. exogenous variables, covariates, input variables, predictor variables or regressors) and a dependent variable (a.k.a. output variable or regress). The major drawbacks of this modeling technique are two-fold: (a) it

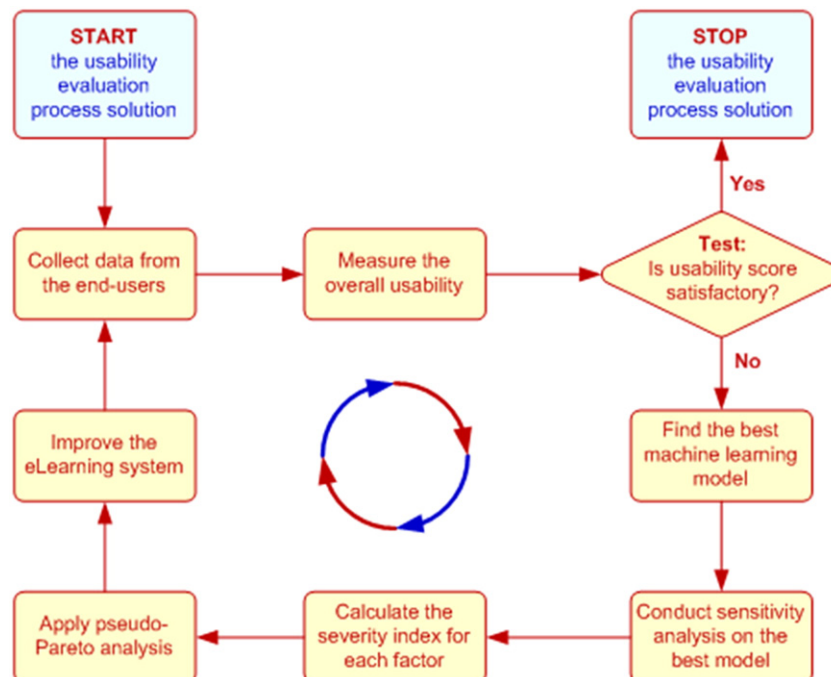


Fig. 1. Proposed eLearning usability evaluation methodology.

**Table 1**  
UseLearn checklist questions/items along with their abbreviated symbols.

Factors and their items	Corresponding UseLearn checklist questions
<i>Error prevention</i>	
EP1	Can multiple but similar tasks be done easily?
EP2	Can the user easily undo selections, actions, errors in arrangement or management of items?
EP3	Do error or warning messages prevent possible errors from occurring?
<i>Visibility</i>	
VIS1	Are options (buttons/selections) logically grouped and labeled?
VIS2	Is the intended functionality clear for each option or selection?
VIS3	Is course content meaningfully arranged with links from the homepage?
<i>Flexibility</i>	
FLEX1	Is the speed of loading course page high enough?
FLEX2	Can users personalize their online learning environment by adding resources, content, learning objects to their own course page?
<i>Course management</i>	
CM1	Does the course contain important information for the online students and link to support areas?
CM2	Does the course provide specific resources to support online student learning?
CM3	Are files easy to upload?
CM4	Are files easy to download and view?
<i>Interactivity, feedback and help</i>	
IFH1	Does the course offer multiple opportunities for interaction and communication among students, to instructor, and to content?
IFH2	Is regular feedback about student performance provided in a timely manner?
IFH3	Is the user provided with sufficient information to know where in the system he/she is?
<i>Accessibility</i>	
ACC1	Are screen features adaptable to individual user preferences?
ACC2	Are accessibility issues addressed throughout the course?
ACC3	Are alternative pathways to course content and activities available?
<i>Consistency and functionality</i>	
CF1	Is consistent form and style used for various titles and headers?
CF2	Do the activity, icon, button, label, and links provide clear purpose/intent that matches the tasks?
CF3	Does the interface provide adequate “back” button functionality to return to a previous screen?
<i>Assessment strategy</i>	
ASST1	Does the eLearning system require students to self-assess their readiness for online instruction prior to class?
ASST2	Are there multiple assessment strategies to measure content knowledge, skills, and performance standards?
ASST3	Are learning objectives, instructional and assessment strategies closely aligned?
<i>Memorability</i>	
MEMO1	Is the user offered sufficient FAQ and human support to obtain necessary help?
MEMO2	Is cognitive load reduced by providing familiarity of items and action sequences?
MEMO3	Is information presented in organized chunks to support learnability and memorability?
MEMO4	Is there sufficient visibility so the user does not have to look for things and try to remember them?
<i>Completeness</i>	
COM1	Are meaningful labels and descriptive links used to support recognition?
COM2	Is the course well organized, easy to navigate, and logical?
COM3	Can you clearly understand all components and structure?
<i>Aesthetics</i>	
AES1	Is there proper use of color or graphics that enhance navigation?
AES2	Are the screens pleasing to look at?
<i>Reducing redundancy</i>	
RED1	Does modifying an action or activity require excessive “redoing” to make a single change?
RED2	Are items visible in multiple places and from multiple paths?
RED3	Are learning objects easily created and reused?

assumes that there is no interaction among the input variables (no collinearity) and (b) it assumes that the relation between the output variable and input variables is linear [74]. *Decision trees* recursively split the data in branches according to a preset criterion (e.g. information gain) to maximize the prediction accuracy resulting in a tree-like structure [62]. To achieve this, they use mathematical algorithms (such as information gain, Gini index, and Chi-squared test) to identify a pair of variables and their thresholds that split the input observation into two or more subgroups [9,34,62,63]. Compared with other machine learning methods, decision trees have the advantage that they are not black box models; hence can easily be explained as rules [22]. *Neural networks* (NNs) have been utilized to model complex relationships among the predictor variables and the dependent variable such as nonlinear functions and multicollinearity [48]. Formally defined, NNs are highly sophisticated analytic techniques capable of predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a process of so-called “learning” from existing data [23]. *Support Vector Machines* (SVMs) are supervised learning methods that generate input–output mapping functions from a set of labeled training data. They belong to a family of generalized linear models which achieve a classification or regression decision based on the value of the linear combination of features. They are also said to belong to the kernel methods [15]. The mapping function in SVMs can be either a classification function (used to categorize the data) or a regression function (used to estimate the numerical value of the desired output).

### 3.2. Model evaluation

#### 3.2.1. Performance criteria

To compare the abovementioned prediction models, two performance criteria are considered: mean squared error (MSE) of the model on testing dataset and correlation between the actual observation for the target variable ( $Y_t$ ) and the predicted value by the model ( $F_t$ ). MSE which is given by the Eq. (2) does not have a rule-of-thumb threshold cut-off value for acceptable models. It is a relative criterion to select the best model, namely the smaller the value the better the model has performed [43].

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - F_t)^2. \quad (2)$$

On the other hand, correlation ( $r_{F_t, Y_t}$ ) which is given by Eq. (3) can be considered as both an absolute measure and a relative measure to determine satisfactory models. For human-related studies, it is recommended that the correlation should be at least 0.3 in behavioral sciences [14] and specifically in usability studies [68]. It would be naïve to express that (unlike the MSE) the higher the correlation the better the performance for the compared models.

$$r_{F_t, Y_t} = \frac{\sum_{i=1}^n (F_{t_i} - \bar{F}_t)(Y_{t_i} - \bar{Y}_t)}{(n-1)(s_{F_t} * s_{Y_t})}. \quad (3)$$

#### 3.2.2. k-Fold cross-validation

In order to minimize the bias associated with the random sampling of the training and holdout data samples in comparing the predictive accuracy of two or more methods, researchers tend to use *k*-fold cross-validation [38]. In *k*-fold cross-validation, also called rotation estimation, the complete dataset ( $D$ ) is randomly split into  $k$  mutually exclusive subsets (the folds:  $D_1, D_2, \dots, D_k$ ) of approximately equal size. The classification model is trained and tested  $k$  times. Each time ( $t \in \{1, 2, \dots, k\}$ ), it is trained on all but one fold ( $D_t$ ) and tested on



the remaining single fold ( $D_t$ ). The cross-validation estimate of the overall performance criteria is calculated as simply the average of the  $k$  individual performance measures as follows,

$$CV = \frac{1}{k} \sum_{i=1}^k PM_i \quad (4)$$

where  $CV$  stands for the cross-validation,  $k$  is the number of folds used, and  $PM$  is the performance measure for each fold [54].

### 3.2.3. Information fusion-based sensitivity analysis

In prediction modeling, there is not a universally accepted “best model” that works for any problem. The best model depends on the scenario being analyzed and the data set being used; and can only be obtained through trial-and-error experimentation [65]. Just as there is not a single best model, there is also not a single best implementation of different model types. Researchers are developing new ways to improve the accuracy and efficiency of prediction models. Therefore, it would be desirable to combine the results developed by multiple models [4]. Use of multiple models should make the forecasts more accurate and more efficient.

Information fusion is the process of intelligently combining the information created and provided by two or more information sources (i.e. models). While there is an on-going debate about the sophistication level of the fusion methods to be employed, there is a general consensus that fusion (combining forecasts and/or predictions) produces more useful information [2]. Combining forecasts can improve accuracy, completeness, and robustness of information, while reducing uncertainty and bias associated with individual models [11]. This multi-model fusion algorithm can be mathematically illustrated as follows [19]:

Given the expected response variable ( $y$ ) and the decision variables ( $x_1, x_2, \dots, x_n$ ) the formulation for any prediction model can be written as Eq. (5)

$$\hat{y} = f(x_1, x_2, \dots, x_n). \quad (5)$$

Prediction model  $f$  can take many forms. For instance, a linear regression model can be written as Eq. (6)

$$f(x_1, x_2, \dots, x_n) = \beta + \sum_{i=1}^n a_i x_i \quad (6)$$

where  $\beta$  is the intercept and  $a_i$ 's are the coefficients for  $x_i$ 's. For a Neural Network model, for a single neuron, it may be written as Eq. (7)

$$f(x_1, x_2, \dots, x_n) = \phi \left( w_0 + \sum_{j=1}^n w_j x_j \right) \quad (7)$$

where  $\phi$  is the transfer function and  $w_i$ 's are the weights for  $x_i$ 's. Given that we use  $m$  number of prediction models, the fusion model can be written as

$$\hat{y}_{fused} = \psi(\hat{y}_{individual,i}) = \psi(f_1(x), f_2(x), \dots, f_m(x)). \quad (8)$$

If  $\psi$  is a linear function, which is the case in this study, then we can write Eq. (9) as

$$\hat{y}_{fused} = \sum_{i=1}^m \omega_i f_i(x) = \omega_1 f_1(x) + \omega_2 f_2(x) + \dots + \omega_m f_m(x) \quad (9)$$

where  $\sum_{i=1}^m \omega_i = 1$ .

The values for  $\omega$ 's are derived from the up-to-now prediction accuracy measure of the individual predictors. That is, the higher the

accuracy of a predictor on independent test cases gets, the larger the weight that is assigned to that predictor type [17].

On the other hand, after determining which prediction models pass the threshold values based on the performance criteria as explained in Section 3.2.1, it is required to determine the rank order for the importance of the independent variables. In artificial neural networks, sensitivity analysis is a method for extracting the cause and effect relationship between the inputs and outputs of a trained ANN model [16,18]. In the process of performing sensitivity analysis, after the model is trained; the ANN learning is disabled so that the network weights are not affected. The fundamental idea is that the sensitivity analysis measures the predictor variables based on the change in modeling performance that occurs if a predictor variable is not included in the model. Hence, the measure of sensitivity of a specific predictor variable is the ratio of the error of the neural network model without the predictor variable to the error of the model that includes this predictor variable [60,61]. The more sensitive the network is to a particular variable, the greater the performance decrease would be in the absence of that variable, and therefore the greater the ratio of importance. The same method is followed in support vector machines to rank the variables in terms of their importance as well according to the sensitivity measure defined as in Eq. (10) [66].

$$S_i = \frac{V_i}{V(F_t)} = \frac{V(E(F_t X_i))}{V(F_t)}. \quad (10)$$

Where  $V(F_t)$  is the unconditional output variance. In the numerator, the expectation operator  $E$  calls for an integral over  $X_{-i}$ ; that is, over all input variables but  $X_i$ , then the variance operator  $V$  implies a further integral over  $X_i$ . Variable importance is then computed as the normalized sensitivity. Saltelli et al. [67] show that Eq. (10) is the proper measure of sensitivity to rank the predictors in order of importance for any combination of interaction and non-orthogonality among predictors. As for the decision trees, variable importance measures were used to judge the relative importance of each predictor variable. Variable importance ranking uses surrogate splitting to produce a scale which is a relative importance measure for each predictor variable included in the analysis.

Considering Eqs. (9) and (10) simultaneously, sensitivity measure of the variable  $n$  with information fused by  $m$  prediction models can then be given by Eq. (11)

$$S_{n(fused)} = \sum_{i=1}^m \omega_i S_{in} = \omega_1 S_{1n} + \omega_2 S_{2n} + \dots + \omega_m S_{mn} \quad (11)$$

where  $\omega$ 's refer to the normalized R-square value of each prediction model with  $m$  models in total and  $S_{in}$  is the sensitivity measure of the  $n$ th variable in the  $i$ th model.

## 4. Case study

### 4.1. Planning and preparing the experiment

In this study, the usability of an eLearning biology course was examined. A cell biology course was selected as the eLearning course tool to be evaluated with the consideration that a qualitative course would be easier to comprehend and followed by the help of an eLearning system rather than a quantitative one.

In creating this online cell biology course, an open area VLE (Virtual Learning Environment) “moodle” was used as the eLearning tool. Moodle® was developed and written by Martin Dougiamas, who was interested in creating a “social constructionist framework” of education within a computerized system. Installation of Moodle® required the source files to be downloaded from its Website, and then to be decompressed onto the local hard disk. Once they were successfully saved, the



Fig. 2. Planning the experiment in Moodle®.

required files were transferred to the web server and then the settings were changed in the configuration files to match our settings at Progress through Training. This installation was aided by several automated pages, which speeded up the process. Apache, PHP, and MySQL/PostgreSQL platforms were used in it [50]. To store any type of data such as user information, courses, tests, test scores and timing information MySQL database software was used along with the programs mentioned above. The opening page of the created cell biology eLearning system appears in Fig. 2.

#### 4.2. Data collection and data preparation for analysis

UseLearn checklist questions were translated into Turkish in order to make them more understandable for the test participants. Therefore, a subsequent preliminary test was performed with 5 students to determine if any translation problems occurred and if any of the questions were confusing to them due to ambiguity in translation. Depending on their reactions and feedback in this pretest, some questions were modified and reworded in favor of clarity. After these modifications, the actual test was applied to collect the data for usability evaluation of the eLearning system in cell biology course.

The test participants consisted of two major parts: 52% were high school students in the 11th grade and 48% were university freshman students majoring in biology. 64% of all participants were male and 36% of them were female. In total 105 students were examined for this study. Proper validation of a usability evaluation study requires the data be gathered from representative users in practical real-life settings [51]. Accurate estimation of usability may necessitate a large sample of users to be surveyed. However, the omnipresent relevance of the data (i.e. even the raw data from a relatively small sample of well-assigned representative users) can be highly informative in identifying the required changes in design [51].

Completing the online biology course and then answering the questions of the checklist took approximately 60–70 min per student. This was a relatively time-consuming study but that was a natural result of

the requirement that the participants had to follow a 30-minute online biology lecture, have an online quiz related to this content, and reply to the UseLearn checklist questions. All tests were carried out in the computer laboratories of a university and a high school by allocating one computer to each participant. The number of the people working around and the noise level were nearly the same as in a classroom environment. The participants received checklist documents containing the aim of the survey, how they would proceed, and some descriptive information such as their names, if they are university or high school student, and their gender.

Preliminary analyses showed that transforming the 5-point Likert-type data points into the [0–1] interval gave much better results in the predictive modeling phase. This transformation was performed using the normalization formula presented as in Eq. (12).

$$\text{New}X = \frac{X - \min X}{\max X - \min X} \quad (12)$$

#### 5. Results and discussion

The average scores of UseLearn checklist items evaluated by the test participants for the aforementioned cell biology eLearning system are summarized in Fig. 3 in ascending order. Note that the existing literature as mentioned in Section 1 claims that the best way to improve the overall usability is to start with the lowest average scored checklist item and continue sequentially. Considering Fig. 3, the conventional usability evaluation strategy would be to start improving the usability problems that are attributed to CM3, CM4, and CM1, respectively.

The sequence of these usability problems would change if our proposed methodology is adopted, which considers not only the average value for the checklist items but also the worth of doing each improvement on the overall usability through the severity index.

In this study, to estimate the performance of the prediction models, a 10-fold cross-validation approach was used. Empirical studies showed that 10 seems to be an optimal number of folds (that optimizes the time

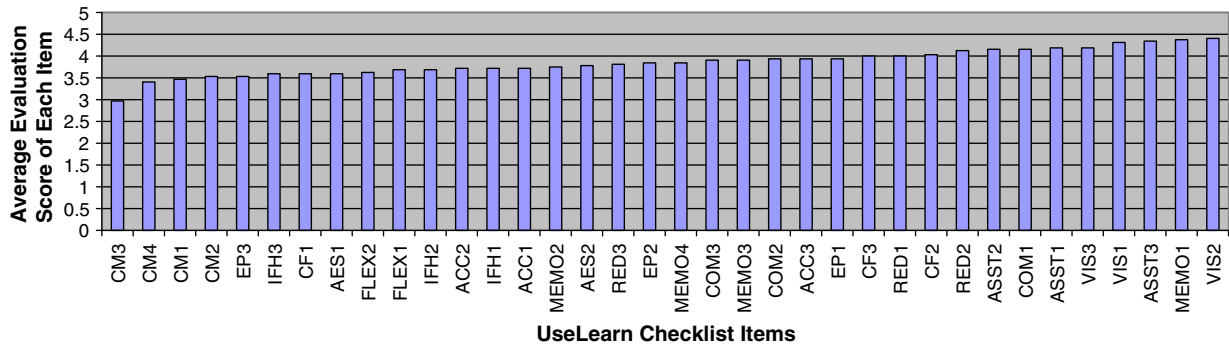


Fig. 3. The average of the UseLearn checklist evaluation scores.

it takes to complete the test while minimizing the bias and variance associated with the validation process) [38]. In 10-fold cross-validation the entire dataset is divided into 10 mutually exclusive subsets (or folds). Each fold is used once to test the performance of the prediction model that is generated from the combined data of the remaining nine folds, leading to 10 independent performance estimates. The prediction models are evaluated based on their MSE and correlation values and the average results for 10-fold cross-validation are summarized in Table 2 where NN, SVM, MLR and, DT refer to neural networks, support vector machines, multiple linear regression, and decision trees, respectively.

Note that the correlations for all the prediction models are well beyond the rule-of-thumb cut-off value of 0.3 [14,68]. The most favorable prediction model that best explains the relationships between the output variable (overall usability) and the input variables in this case study is the multi-layer perceptron (MLP) neural network. The MLP is known to be a powerful and robust function approximator for prediction and classification problems. It is arguably the most commonly used and well-studied NN architecture [26]. In our experimental runs it was also observed that MLP performs better than other machine learning methods. In fact, Hornik et al. [26] empirically showed that given the right size and the structure, MLP is capable of learning arbitrarily complex nonlinear functions to arbitrary accuracy levels. The MLP is essentially the collection of nonlinear neurons (a.k.a. perceptrons) organized and connected to each other in a feed-forward multi-layer structure.

Table 2  
10-fold cross-validated model results.

Predictive models	Performance measures	
	MSE	Correlation
<b>NN</b>		
Dynamic	0.072	0.770
MLP	0.068	0.786
Prune	0.075	0.780
RBF	0.070	0.668
<b>SVM</b>		
Linear	0.122	0.514
Polynomial	0.101	0.752
RBF	0.085	0.757
<b>MLR</b>		
Backward	0.100	0.615
Forward	0.076	0.615
Stepwise	0.066	0.633
<b>DT</b>		
CART	0.092	0.771
CHAID	0.090	0.741

Fig. 4 is a graphical representation of the MLP-type neural network model which gave the best MSE and correlation values in this study. It has 37 neurons (processing elements-PE) in the input layer (referring to 37 checklist items), 19 and 10 neurons in the first and the second hidden layers, respectively. The output layer has one neuron which refers to the output variable (overall usability).

Using the information fusion-based sensitivity analysis scores of each checklist item as in Eq. (11) via each model listed in Table 2 and the reciprocal of the mean values for each from Fig. 3, severity index for each item was calculated as explained in Eq. (1). A subsequent step in the proposed methodology is to employ a Pareto analysis-like procedure to select “vital few” variables from “trivial many” [57]. By applying a pseudo-Pareto analysis to the checklist items, we ranked them in descending order as in Fig. 5. It represents the severity indices for each checklist item. Note that the sequence of the items to be improved is different than that in Fig. 3. The severity of the predictor variables was extracted by a pseudo-Pareto analysis in a similar fashion to regular Pareto rule with a small difference: instead of sticking to conventional 80/20 rule as suggested by Pareto

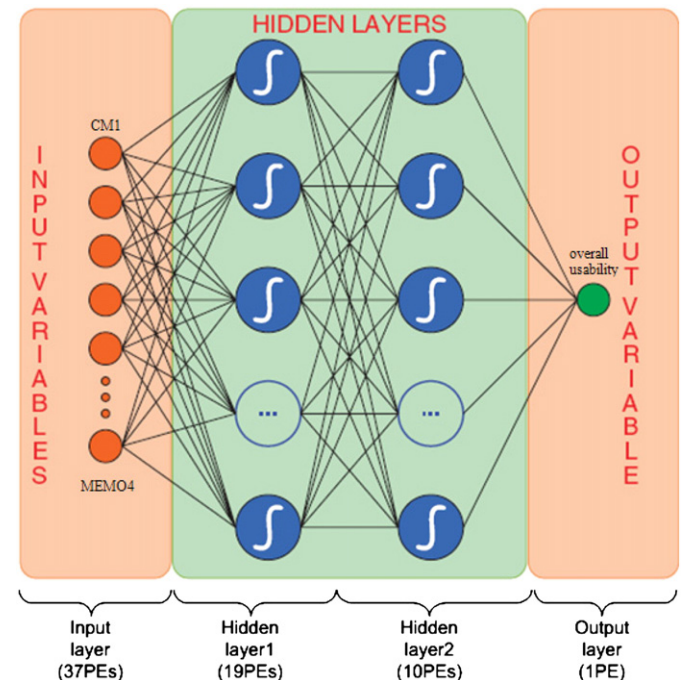


Fig. 4. MLP architecture as the best prediction model.

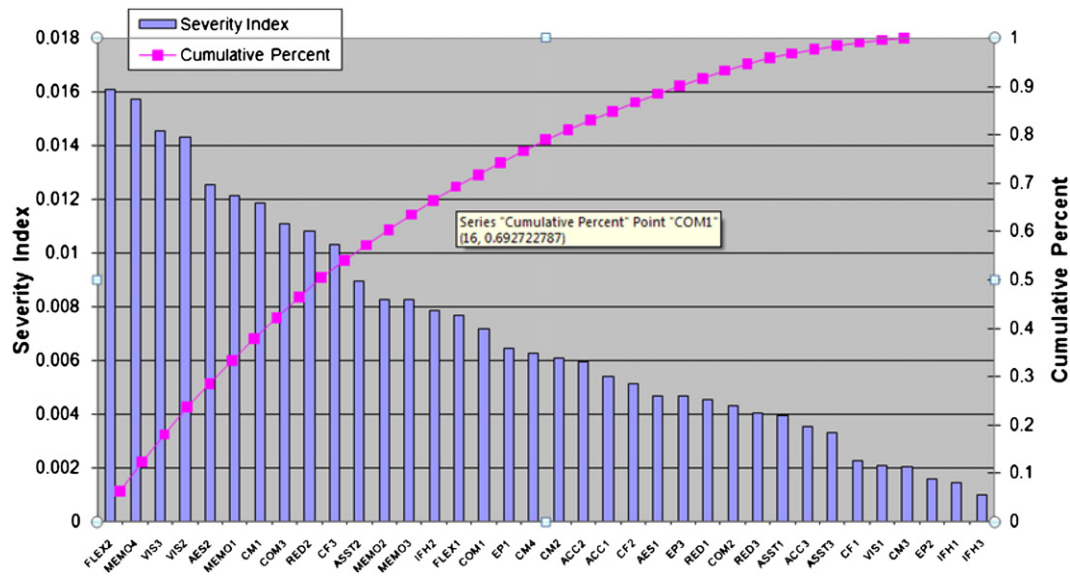


Fig. 5. Pareto chart for severity indices of the checklist items (input variables).

[57], in this case study we propose to deal with 43% of the causes (16 of the all 37 checklist items) which corresponds to almost 70% of the all usability-related problems as shown in Fig. 5. This means the usability expert would start improving the usability problems that are attributed to FLEX2, MEMO4... up to COM1 due to the results received by the Pareto chart in Fig. 5.

Usually usability experts do not have enough time to deal with all of the problematic usability dimensions to improve the overall eLearning system usability. Therefore, they would like to know which measure(s) is/are the most critical one(s) in improving the usability. However, the general tendency in the literature as summarized in Section 1 shows

that most of the usability evaluation methods adopt a straightforward approach by considering only the most problematic measures with the smallest mean values (as ranked in Fig. 3). This conventional approach lacks one critical point of view: Is it really worth confronting that/those measure(s)? In other words, although the measure has the smallest mean value based on a survey-based evaluation; would it make a big effect on the overall usability index after improving the corresponding checklist item accordingly? Therefore, our methodology proposes to take into account both Fig. 3 mean values for each checklist item and also sensitivity scores of these measures on the overall usability. This simultaneous consideration would help determine which

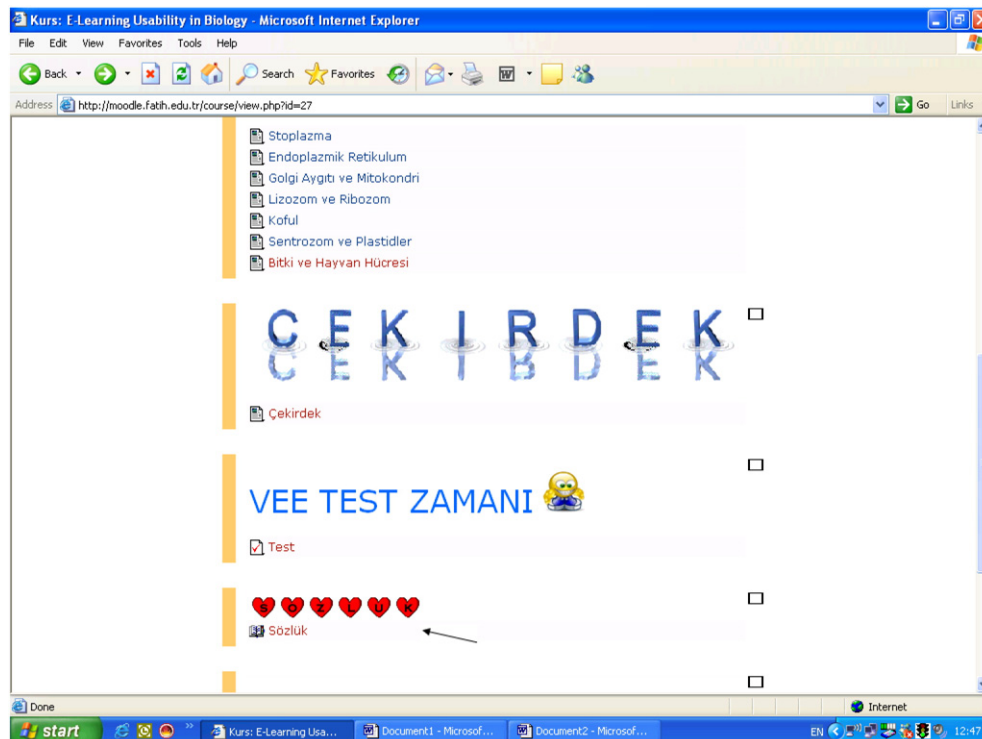


Fig. 6. A sample screenshot for usability improvement through MEMO4 item.



worthy measures to focus on for further improvement sequentially as shown in Fig. 5. It basically requires that checklist items with the smallest mean values be chosen provided that a unit change in them would make a significant change on the usability. The focus here would be our combined measure, namely the severity index, which considers both perspectives: the smallest checklist item along with the biggest impact. **If only Fig. 3 results are taken into account as in the conventional usability evaluation approaches, the top 5 problematic measures would be CM3, CM4, CM1, CM2, and EP3.**

**Fig. 5 provides a modified importance sequence of checklist items by considering their impact on the overall usability as well.** Note that the sequence of the top 5 most important measures is different than the previously selected ones through Fig. 3. This modification indicates that in order to improve the usability of the eLearning system on hand the most critical and problematic items to focus on are FLEX2, MEMO4, VIS3, VIS2, and AES2, respectively due to their higher contribution to usability improvement. To exemplify the usability problems that were detected by the proposed method we present the FLEX2 item which refers to the UseLearn checklist question “*Can users personalize their online learning environment by adding resources, content, learning objects to their own course page?*” **Before the analysis was conducted, the eLearning system did not provide such flexibility for the students.** As for the next usability problem, it is defined by MEMO4 which refers to the UseLearn checklist question “*Is there sufficient visibility so the user does not have to look for things and try to remember them?*”

In the initial design of the eLearning system, it did not have a dictionary to provide short descriptions of the terminology used in the biology course. After this detection, we added such a dictionary for the ease of future end-users/students. This change is illustrated in Fig. 6. Note that “dictionary” means “sozluk” in Turkish language.

## 6. Conclusions and further research direction

A machine learning-based approach for usability evaluation of eLearning systems was proposed and implemented by the help of an online cell biology course using the eLearning system, Moodle®. The analysis of experimental results showed that the proposed methodology supports the identification of usability problems and definition of relevant improvement strategies.

The main advantage of the proposed method is the selection of the most important checklist items considering their contribution to the overall usability based on a newly created metric (i.e. the severity index) and thus optimally utilizing the time and effort for usability improvement. Therefore, it can be claimed that this is the foremost quantitative method among the usability evaluation methods because the results of the case study showed that this algorithm successfully reveals the significant usability problems in order to improve the eLearning usability. This is achieved by merely focusing on some, not all, of the usability problems in a sequential sense. The study also presents uniqueness in that it is the foremost one which would help the usability experts decide where to stop while improving the usability problems of a system through pseudo-Pareto analysis. The limited resources of the eLearning usability experts (e.g. time and money) can be allocated only for the severe checklist items which have crucial impacts on its usability. The case study results validate the proposed method and shows that it is flexible enough to be applied to other eLearning systems.

Further research efforts may incorporate the application of structural equation modeling (SEM) or partial least squares (PLS) techniques because they provide a clearer cause-and-effect relationship among the input variables and the output variable. Additionally, they can handle the effects of both the observed/measured and the corresponding latent (unobserved/unmeasured) variables on the usability in a step-by-step fashion; hence provide more granularity in the analysis. For example, a usability expert might be interested in first determining

which of the checklist factors in Table 1 are more critical than the others as a high-level measure (error prevention, visibility, flexibility, etc.). Only then s/he might be willing to go further in depth to determine the exact cause of the low usability (for example EP1, VIS2 and FLEX2). The concern related to the potential nonlinearity relationships among the output and input variables that lies in those methods (SEM and PLS) can also be overcome by deploying the nonlinear partial least squares (NLPLS) algorithm which is an extension of regular PLS with neural networks. This would make the algorithm as powerful as the machine learning methods that can capture the nonlinearity while simultaneously explaining the cause-and-effect relationships at a desired granularity as mentioned above.

## References

- [1] J.B. Arbaugh, R. Benbunan-Fich, The importance of participant interaction in online environments, *Decision Support Systems* 43 (2007) 853–865.
- [2] J.S. Armstrong, Combining forecasts, in: J.S. Armstrong (Ed.), *Principles of Forecasting*, Kluwer Academic Publishers, Norwell, 2002, pp. 418–439.
- [3] E.M. Babiker, H. Fujihara, C.D.B. Boyle, A metric for hypertext usability, *Proceedings of 11th Annual International Conference on Systems Documentation*, ACM Press, 1991, pp. 95–104.
- [4] R. Batchelor, P. Dua, Forecaster diversity and the benefits of combining forecasts, *Management Science* 41 (1995) 68–75.
- [5] M. Ben Ayed, H. Ltfi, C. Kolski, A.M. Alimi, A user-centered approach for the design and implementation of KDD-based DSS: a case study in the healthcare domain, *Decision Support Systems* 50 (2010) 64–78.
- [6] N. Bevan, Measuring usability as quality of use, *Software Quality Journal* 4 (1995) 115–140.
- [7] N. Bevan, Quality in use: meeting user needs for quality, *Journal of Systems and Software* 49 (1999) 89–96.
- [8] P. Bharat, A. Chaudhury, An empirical investigation of decision-making satisfaction in web-based decision support systems, *Decision Support Systems* 37 (2004) 187–197.
- [9] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees*, Wadsworth & Brooks/Cole Advanced Books, Monterey, Ca., 1984.
- [10] J. Brooke, SUS: a “Quick and Dirty” usability scale, in: P. Jordan, B. Thomas, B. Weerdmeester (Eds.), *Usability Evaluation in Industry*, Taylor and Francis, London, 1996, pp. 189–194.
- [11] C.W. Chase Jr., Composite forecasting: combining forecasts for improved accuracy, *Journal of Business Forecasting Methods & Systems* 19 (2000) 2–22.
- [12] J.P. Chin, V.A. Diehl, K.L. Norman, Development of an instrument measuring user satisfaction of the human–computer interface, *Proceedings of CHI’88*, ACM Press, Washington, D.C., 1988, pp. 213–218.
- [13] C.M. Chiu, M.H. Hsu, S.Y. Sun, T.C. Lin, P.C. Sun, Usability, quality, value and E-learning continuance decisions, *Computers in Education* 45 (2005) 399–416.
- [14] J. Cohen, P. Cohen, S.G. West, L.S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, L. Erlbaum Associates Mahwah, New Jersey, 2003.
- [15] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, London, 2000.
- [16] G. Davis, Sensitivity analysis in neural net solutions, *IEEE Transactions on Systems, Man, and Cybernetics* 19 (1989) 1078–1082.
- [17] D. Delen, S. Hawamdeh, A holistic framework for knowledge discovery and management, *Communications of the ACM* 52 (2009) 141–145.
- [18] D. Delen, E. Sirakaya, Determining the efficacy of data-mining methods in predicting gaming ballot outcomes, *Journal of Hospitality and Tourism Research* 30 (3) (2006) 313–332.
- [19] D. Delen, R. Sharda, P. Kumar, Movie forecast guru: a web-based DSS for Hollywood managers, *Decision Support Systems* 43 (2007) 1151–1170.
- [20] S. Downey, R.M. Wentling, T. Wentling, A. Wadsworth, The relationship between national culture and the usability of an E-learning system, *Human Resource Development International* 8 (2005) 47–64.
- [21] B. Efron, R. Tibshirani, Statistical data analysis in the computer age, *Science* 253 (1991) 390–395.
- [22] X. Fang, C.W. Holsapple, An empirical study of web site navigation structures’ impacts on web site usability, *Decision Support Systems* 43 (2007) 476–491.
- [23] S. Haykin, *Neural Networks and Learning Machines*, 3rd edition Prentice Hall, New Jersey, 2008.
- [24] A. Hope, P. Guiton, *Strategies for Sustainable Open and Distance Learning*, Routledge Taylor and Francis Group, London, 2006.
- [25] K. Hornbaek, Current practice in measuring usability: challenges to studies and research, *International Journal of Human Computer Studies* 64 (2006) 79–102.
- [26] K. Hornik, M. Stinchcombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feed-forward network, *Neural Networks* 3 (1990) 359–366.
- [27] C.M. Hsu, Y.C. Yeh, J. Yen, Development of design criteria and evaluation scale for web-based learning platforms, *International Journal of Industrial Ergonomics* 39 (2009) 90–95.

- [28] P.J.-H. Hu, W. Hui, Examining the role of learning engagement in technology-mediated learning and its effects on learning effectiveness and satisfaction, *Decision Support Systems* 53 (2012) 782–792.
- [29] IEEE, IEEE Standard 161: Standard for a Software Quality Metrics Methodology, 1998.
- [30] ISO, ISO 9241–11, 98: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTS)–Part 11: Guidance on Usability, 1998.
- [31] ISO, ISO 13407, 99: Human-Centered Design Processes for Interactive Systems, 1999.
- [32] ISO/IEC, ISO 9126: Information Technology – Software Quality Characteristics and Metrics, 1991.
- [33] N. Juristo, A.M. Moreno, M.I. Sanchez-Segura, Analyzing the impact of usability on software design, *Journal of Systems and Software* 80 (2007) 1506–1516.
- [34] G.V. Kass, An exploratory technique for investigating large quantities of categorical data, *Applied Statistics* 29 (1980) 119–127.
- [35] B. Keevil, Measuring the usability index of your web site, *Proceedings of the 16th Annual International Conference on Computer Documentation, SIGDOC*, ACM Press, 1998, pp. 271–277.
- [36] J. Kirakowski, The Software Usability Measurement Inventory (Sumi): background and usage, in: P. Jordan, B. Thomas, B. Weerdmeester (Eds.), *Usability Evaluation in Industry*, Taylor and Francis, London, 1996, pp. 169–178.
- [37] J. Kirakowski, M. Corbett, SUMI: the software measurement inventory, *British Journal of Educational Technology* 24 (1993) 210–212.
- [38] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proceedings of the 14th International Conference on AI (IJCAI)*, Morgan Kaufmann, San Mateo, CA, 1995, pp. 1137–1145.
- [39] Y. Lee, K.A. Kozar, Understanding of website usability: specifying and measuring constructs and their relationships, *Decision Support Systems* 52 (2012) 450–463.
- [40] J.R. Lewis, Psychometric evaluation of the PSSUQ using data from five years of usability studies, *International Journal of Human Computer Interaction* 14 (2002) 463–488.
- [41] H.X. Lin, Y. Choong, G. Salvendy, A proposed index of usability: a method for comparing the relative usability of different software systems, *Behavior and Information Technology* 16 (1997) 267–278.
- [42] M. Macleod, R. Rengger, The development of drum: a software tool for video-assisted usability evaluation of hypermedia, *Proceedings of The HCI Conference on People and Computers VIII*, Cambridge University Press, Loughborough, UK, 1993, pp. 293–309.
- [43] S. Makridakis, S.C. Wheelwright, R.J. Hyndman, *Forecasting: Methods and Applications*, John Wiley and Sons, New York, 1998.
- [44] M. Matera, M.F. Costabile, F. Garzotto, P. Paolini, SUE inspection: an effective method for systematic usability evaluation of hypermedia, *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* 32 (2002) 93–103.
- [45] M. Mcgee, Master usability scaling: magnitude estimation and master scaling applied to usability measurement, *Proceedings of CHI 2004*, ACM Press, Washington, DC, 2004, pp. 335–342.
- [46] P. McNamara, Usability and Elearning, Final Coursework Essay: Certificate in Online Education and Training, Institute of Education, University of London. Available at: <http://Nte.Unifr.Ch/Img/Pdf/Pmcmnarausabilityelearning.Pdf> 2003.
- [47] M.J. Miller, Usability in Elearning, ASTD's Source for Elearning, Learning Circuit. Available At: [http://Www.Astd.Org/Lc/2005/0105\\_Miller.Html](http://Www.Astd.Org/Lc/2005/0105_Miller.Html) 2005.
- [48] T. Mitchell, *Machine Learning*, McGraw-Hill, New York, 1997.
- [49] R. Molich, J. Nielsen, Improving a human–computer dialogue, *Communications of the ACM* 33 (1990) 338–348.
- [50] Moodle, A Free Open Source Course Management System for Online Learning, 2011. ([www.Moodle.Org](http://www.Moodle.Org) and <http://Moodle.Fatih.Edu.Tr>).
- [51] J. Nielsen, *Usability Engineering*, Academic Press Professional, Boston, 1993.
- [52] J. Nielsen, R.L. Mack, *Usability Inspection Methods*, John Wiley and Sons, New York, 1994.
- [53] K. Norman, B. Schneiderman, Questionnaire for User Interaction Satisfaction, HCI Lab, College Park, University of Maryland, 1989.
- [54] D.L. Olson, D. Delen, *Advanced Data Mining Techniques*, Springer, New York, 2008.
- [55] A. Oztekin, Z.J. Kong, O. Uysal, UseLearn: a novel checklist and usability evaluation method for eLearning systems by criticality metric analysis, *International Journal of Industrial Ergonomics* 40 (2010) 455–469.
- [56] A. Oztekin, A. Nikov, S. Zaim, UWIS: an assessment methodology for usability of web-based information systems, *Journal of Systems and Software* 82 (2009) 2038–2050.
- [57] V. Pareto, *Manuale D'economia Politica*. (English Translation by A.M. Kelly), 1971.
- [58] O. Parlangeli, E. Marchigiani, S. Bagnara, Multimedia systems in distance education: effects of usability on learning, *Interacting with Computers* 12 (1999) 37–49.
- [59] K. Pearson, A. Lee, On the generalized probable error in multiple normal correlation, *Biometrika* 6 (1908) 59–68.
- [60] J.C. Principe, D. Xu, J.W. Fisher III, *Information-Theoretic Learning*, in *Unsupervised Adaptive Filtering*, John Wiley and Sons, New York, 2000.
- [61] J.C. Principe, N.R. Euliano, W.C. Lefebvre, *Neural and Adaptive Systems*, John Wiley and Sons, 2001.
- [62] J. Quinlan, Induction of decision trees, *Machine Learning* 1 (1986) 81–106.
- [63] J. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, 1993.
- [64] R. Rengger, M. Macleod, R. Bowden, A. Drynan, M. Blaney, *Musical Performance Measurement Handbook*, National Physical Laboratory, Dtc, Teddington, UK, 1993.
- [65] E. Ruiz, F.H. Mieto, A note on linear combination of predictors, *Statistics & Probability Letters* 47 (2000) 351–356.
- [66] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications* 145 (2002) 280–297.
- [67] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, *Sensitivity Analysis in Practice – A Guide to Assessing Scientific Models*, John Wiley and Sons, 2004.
- [68] J. Sauro, J.R. Lewis, Correlations among prototypical usability metrics: evidence for the construct of usability, *Proceedings of The 27th International Conference on Human Factors in Computing Systems*, Association for Computing Machinery, Boston, MA, 2009, pp. 1609–1618.
- [69] B. Schneiderman, *Designing the User Interface: Strategies for Effective Human–Computer Interaction*, Addison Wesley, Menlo Park, CA, 1998.
- [70] A. Seffah, T. Mohamed, H. Habieb-Mammar, A. Abran, Reconciling usability and interactive system architecture using patterns, *Journal of Systems and Software* 81 (2008) 1845–1852.
- [71] D. Squires, J. Preece, Predicting quality in educational software: evaluating for learning, usability, and the synergy between them, *Interacting with Computers* 11 (1999) 467–483.
- [72] M.A. Storey, B. Phillips, M. Maczewski, M. Wang, Evaluating the usability of web-based learning tools, *Educational Technology & Society* 5 (2002) 91–103.
- [73] P.-C. Sun, H.K. Cheng, G. Finger, Critical functionalities of a successful e-learning system – an analysis from instructors' cognitive structure toward system usage, *Decision Support Systems* 48 (2009) 293–302.
- [74] S. Weisberg, *Applied Linear Regression*, John Wiley and Sons, New York, 1980.



**Dr. Asil Oztekin** is an Assistant Professor of Operations and Information Systems Department, Manning School of Business at University of Massachusetts Lowell. He received the B.S. degree from Yildiz Technical University, Istanbul, Turkey, in 2004, and the M.S. degree from Fatih University, Istanbul, Turkey, in 2006, both in industrial engineering. He also completed the Ph.D. degree in the School of Industrial Engineering and Management at Oklahoma State University (OSU), Stillwater, in 2010. Prior to joining UMass Lowell, he worked as a Visiting Assistant Professor in the Department of Statistics at Oklahoma State University. His research interests include human–computer interaction, medical informatics, healthcare engineering, decision analysis, multivariate data analysis, data mining, and quality engineering. His work

has been published in leading journals such as *Decision Support Systems*, *International Journal of Production Research*, *Production Planning & Control*, *International Journal of Medical Informatics*, *Artificial Intelligence in Medicine* and *International Journal of Industrial Ergonomics*. He edited a special issue as a guest editor entitled “Intelligent Computational Techniques in Science, Engineering, and Business” in *Expert Systems with Applications* journal. Dr. Oztekin is serving as an editorial review board member of the for the *Journal of Computer Information Systems*, *International Journal of Services and Operations Management*, *International Journal of Operations Research and Information Systems*, *Journal of Modelling in Management*, *International Journal of Data Analysis Techniques & Strategies*, *International Journal of Business Analytics*, and for *International Journal of Business Intelligence and Systems Engineering*. His research in medical decision making has been recently funded by the *Advancing Research, Scholarship, and Creative Work Grant* at UMass Lowell. Dr. Oztekin has served as the session chair for the mini-tracks “Predictive Analytics and Big Data” at the HICSS 46 and “Data, Text, and Web Mining for Managerial Decision Support” at the HICSS 47. He is a member of HIMMS, ASQ, IIE, and INFORMS and was the recipient of the Alpha Pi Mu Outstanding Industrial Engineering and Management Research Assistant Award from OSU in 2009.



**Dr. Dursun Delen** is the holder of William S. Spears and Neal Patterson Endowed Chairs in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University (OSU). He received his Ph.D. in Industrial Engineering and Management from OSU in 1997. Prior to his appointment as an Assistant Professor at OSU in 2001, he worked for a privately-owned research and consultancy company, Knowledge Based Systems Inc., in College Station, Texas, as a research scientist for five years, during which he led a number of decision support and other information systems related research projects funded by federal agencies, including DoD, NASA, NIST and DOE. His research has appeared in major journals

including *Decision Support Systems*, *Communications of the ACM*, *Computers and Operations Research*, *Computers in Industry*, *Journal of Production Operations Management*, *Artificial Intelligence in Medicine*, *Expert Systems with Applications*, among others. He recently published four books: *Advanced Data Mining Techniques* with Springer, 2008; *Decision Support and Business Intelligence Systems* with Prentice Hall, 2010; *Business Intelligence: A Managerial Approach*, with Prentice Hall, 2010; and *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, with Elsevier, 2012. He is often invited to national and international conferences for keynote addresses on topics related to Data/Text Mining, Business Intelligence, Decision Support Systems, and Knowledge Management. He served as the general co-chair for the 4th International Conference on Network Computing and Advanced Information Management (September 2–4, 2008 in Seoul, South Korea), and regularly chairs tracks and mini-tracks at various information system conferences. He is the associate editor-in-chief for *International Journal of Experimental Algorithms*, associate editor for *International Journal of RF Technologies*, and is on editorial boards of five other technical journals. His research and teaching interests are in data and text mining, decision support systems, knowledge management, business intelligence and enterprise modeling.



**Dr. Ali Turkyilmaz** is the head of Industrial Engineering Department at Fatih University. He got his Ph.D. degree in Industrial Engineering Department of Istanbul Technical University. His main research interests are production and service sector management, total quality management, applied statistics, and decision making. He is also a consultant and trainer for industrial companies, an independent expert for EU projects, and a member of Turkish Quality Organization and Turkish Operations Research Foundation. Dr. Turkyilmaz is serving on the editorial review board of *Industrial Management & Data Systems*, *International Journal of Service and Standards*, and *International Journal of Synergy & Research*.



**Dr. Selim Zaim** has received his B.S. degree in Mechanical Engineering from Istanbul Technical University and his Ph.D. degree in Production and Operations Management from Istanbul University. Dr. Zaim has been serving as a Professor in the College of Technology at Marmara University. Dr. Zaim has published over 100 articles in various journals and congress proceedings. His current scholarly interests focus on multivariate data analysis, human–computer interaction, and multi-criteria decision making. Dr. Zaim reviews papers for a variety of journals. He is a member of Industrial Management and Development Association (IMDA) and Quality Association in Turkey (KALDER).