Evaluation Cycle Management - Model for Selection of the most Applicable Learning Management System

MATIJA PIPAN, TANJA ARH, BORKA JERMAN BLAZIC
Laboratory for Open Systems and Networks
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana
SLOVENIA
{matic, tanja, borka}@e5.ijs.si http://www.e5.ijs.si

Abstract: - The paper deals with a complex decision-making problem, the evaluation and selection of most applicable Learning Management Systems (LMS) in which several objectives - referring to the definite group of users - like social, technical, environmental, and economic impacts, must be simultaneously taken into account. The general aim of our case study was focusing on the usability and applicability aspects of LMSs in relation to definite target group and users: employees in the Slovenian SMEs with a basic knowledge of ICT. We introduce Evaluation Cycle Management (ECM), a support methodology aimed at the evaluation of options that occur in the decision-making processes. ECM is based on Multi-attribute decision making (Criteria Evaluation) and Usability Testing (Usability Evaluation). The Multi-attribute decision making in the first phase of ECM presents an approach to the development of a qualitative hierarchical decision model that is based on DEX, an expert system shell for multi-attribute decision support. The second phase of ECM is aimed at Usability Testing on end users.

Key-Words: - Innovative E-learning Software, Education, Learning Management Systems (LMS), Evaluation Cycle Management (ECM), Criteria Evaluation, Usability Testing, Multi Attribute Decision Making

1 Introduction

Considering the abundance of e-learning systems that have offered education over the Internet during the past decade, it is not surprising that there has been growing interest in identifying design principles and features that can enhance user satisfaction. User satisfaction with technologies related to distance and collaborative learning applications has been found to be significantly associated with usability, that is, the effectiveness, efficiency and satisfaction that it gives to its user in a given context of use and task. The usability of an educational environment is related to its pedagogical value and evaluation of its usability is part of the processes of establishing its quality. In the literature, there are numerous recommendations for the design of pages, text, graphics, and navigation in Learning Management Systems (LMSs), but in spite of that, it is still recognized that "severe usability problems are present and common". However, despite the increased awareness of these problems when adopting internet-based education the usability of e-learning systems has still not been sufficiently explored and solutions not yet provided.

These are some of the realizations that led us to perform the case study described in this paper and to analyze the results. The case study was undertaken as part of an EU project centered on the issues of introducing internet-based education in a region that suffers from a low level of business-oriented usage of the Internet and related e-services together with a relatively high level of unemployment. We found the environment and the context of this study extremely suitable for an evaluation and assessment of the usability of the Learning Management Systems, and to try to identify the "threshold of acceptability beyond which users can begin to interact productively and voluntarily instead of simply acting and reacting".

The usability of a Learning Management System is often perceived to be the province of the technical expert rather than the content expert; however, technical knowledge is insufficient when it comes to designing and testing systems intended for e-learning. This paper presents one attempt to apply and evaluate different Learning Management Systems and also to discuss the findings in an assessment of the *learnability*, *effectiveness, efficiency* and level of *satisfaction* of an LMS. Results of the case study can provide a better understanding of the ECM methodology, development of multi-attribute decision making and usability testing.

To achieve the proposed objectives, we will present and describe *Evaluation Cycle Management (ECM)*, a novel methodology aimed at the evaluation of options that

occur in decision-making processes. This section presents a brief introduction to the *Multi-attribute decision making (Criteria evaluation)* as a first phase of ECM and *Usability evaluation* as the second phase of ECM. Each of the applied methods is followed by a description of the scenario and a study of the results. Finally, the paper ends with a discussion of the overall results, conclusions and implications [3,8].

2 Evaluation Cycle Management (ECM)

2.1 What is ECM?

To assure that a product is good enough to satisfy all the needs and requirements of the users and other potential stakeholders, such as the users' clients or managers, we need to verify the products' characteristics and assess its acceptability within various categories. Several unique methods and techniques for evaluating products/systems are known, as well as many possible ways of combining various evaluation methods.

Evaluation Cycle Management (ECM), which was developed by the authors, can be classified as a combined evaluation system, because it is composed of two independent evaluation methods: Multi-attribute decision making (Criteria evaluation) and Usability testing (Usability evaluation).

2.2 Architecture of ECM

The principal feature that characterizes the Evaluation Cycle Management (ECM) is a two-phase evaluation method with a feedback loop. The first phase of ECM includes Multi-attribute decision making and the second, Usability evaluation. The results gained from the Multi-attribute decision making model (first phase), developed by experts, is being verified on users as well. In case user usability testing (second phase) shows overly significant changes between the presupposed and the gained results, we return to the first phase and correct the multi-attribute model on the basis of the analysis results. When an observed product/system gains good results with the user testing, or only minor corrections are needed, such a system is recommended.

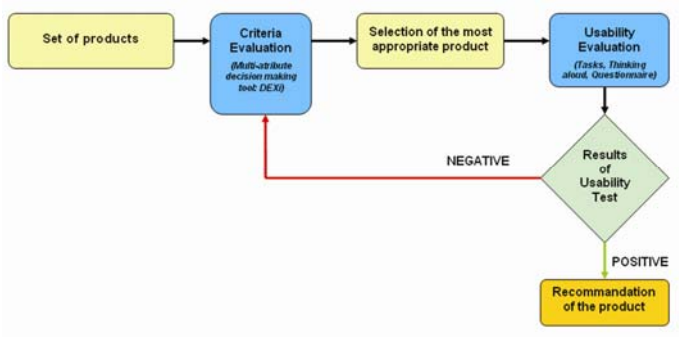

Fig.1: Architecture of ECM

The key advantage of the ECM methodology as seen by the authors is that in the first phase of evaluation (evaluation using a multi-attribute decision making model) only one – the most suitable solution – is chosen which leads to lowered costs and decreased use of time regarding continued evaluation in the second phase (usability testing), for only one solution is subjected to testing and not all [6].

a) The First Phase of ECM: *Criteria Evaluation*
Decision making is a process of selecting a particular option from a set of possibilities, so as to best satisfy the aims or goals of the decision maker. In practice, the *options* (also called *alternatives*) are objects or actions of (approximately) the same type, such as different computer systems, different people applying for a particular job, different investment strategies, and different e-learning technologies. A number of methods and computer-based systems have been developed. They are mainly studied in the framework of decision support systems, operations research and management sciences, and decision theory or decision analysis.

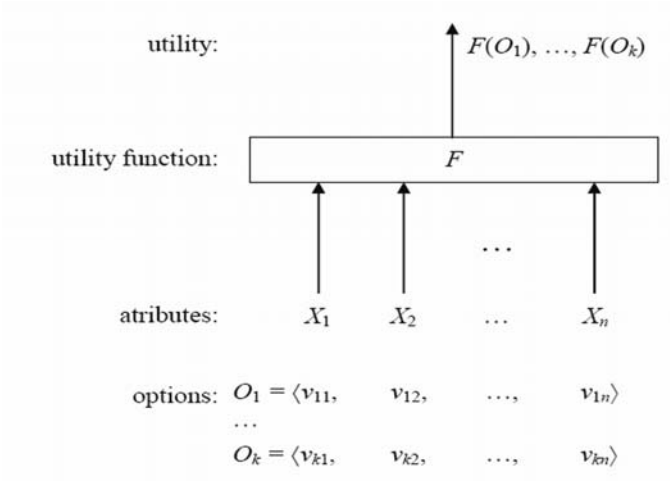

Fig.2: General concept of multi-attribute decision making

The contribution to these fields has been the development of an expert system shell for multi-

attribute decision support DEX. DEX itself is designed as an interactive *expert system shell* that provides tools for building and verifying a knowledge base, evaluating options and explaining the results. The structure of the knowledge base and evaluation procedures closely correspond to the multi-attribute decision making paradigm. This makes the system specialized for decision support [4].

Some recent developments have made the hierarchical decision model approach very attractive also for problems in web based education and e-learning. In particular, some newly developed methods, including DEX, facilitate the design of qualitative (or symbolic) decision models. In contrast to traditional quantitative (numeric) models, the qualitative ones use symbolic variables. These seem to be better suited for dealing with 'soft' decision problems, which are typical for education and e-learning: less structured and less formalized problems that involve a great deal of expert judgments as opposed to exact formal modeling and computation. In next section we present the approach to the development and application of qualitative hierarchical decision models that is based on the DEX shell [2,3].

b) The Second Phase of ECM: *Usability Evaluation*
Usability is most often defined as the ease of use and acceptability of a system for a particular class of users carrying out specific tasks in a specific environment. Ease of use affects the users' performance and their satisfaction, while acceptability affects whether the system or product is used.

In the case of Evaluation Cycle Management, only the usability evaluation on end users reveals the real value of the observed system/product, which has been chosen on the first evaluation phase with the aid of the multi-attribute decision making technique. Testing with end users is one of the most fundamental usability methods and one which is irreplaceable in the second phase of the ECM methodology, because it provides direct information about how people use the system/product, what their exact problems are with the concrete interface and it verifies its expected usefulness for the system user.

In usability testing it is very important to pay attention to the issues of reliability and validity. Reliability is the question of whether one would get the same results if the test were to be repeated, and validity is the question of whether the result actually reflects the usability issues one wants to test. To ensure reliability and validity of the usability evaluation outcome, several points need to be considered: representativeness of test participants for real users of the system, realness of testing tasks as compared to actual tasks performed by real users, accuracy of observations of test participants behaviors, and sensitivity of measuring tools.

There are several unique methods and techniques for testing usability as well as many possible ways of combining various usability methods. A combination of methods and techniques used in the second phase of ECM usually contains: task scenarios, think aloud, field observations, questionnaires and participant debriefing.

3 Selection of the most appropriate LMS

3.1 Identification of the Problem

Parallel to the wide range of possibilities offered by new generations of educational technologies, a number of Learning Management Systems (LMSs) which support e-learning have been developed and are available on the market. Consequently, customers are often faced with the dilemma of how to choose the optimum LMS for the implementation of the education process for a definite target group. The general aim of our case study was focusing on the usability and applicability aspects of LMSs in relation to definite target group and users: employees in the Slovenian Drava-Mura Region SMEs with a basic knowledge of ICT.

3.2 Criteria Evaluation - Development of the Multi-Attribute Decision Making Model

The decision-making process was divided into four phases: *(1) criteria identification and criteria structuring, (2) utility function definition (decision rules), (3) description of variants, (4) LMS evaluation and analysis* [1,2,5].

- *Identification, Description and Criteria Structuring*

This phase provides descriptions of criteria which are the components of the decision-making model. The criteria are divided into three main scopes: *Student's learning environment, System, technology & standards and Tutoring & didactics.* These three scopes represent the skeleton of the multi-attribute model. The criteria can include the following values: 'low', 'average' or 'high'; the only exception being the criteria where it is impossible to determine an intermediate value. All values have an increasing range (low value is worst than high value).

- *Utility Function*

The *tree of criteria* defines the structure of the evaluation model by defining the criteria and their interdependence. In the final outcome, this means that the overall evaluation of the LMS depends on 57 criteria. On the other hand, the criteria tree does not define the aggregation, i.e., the procedure that combines the values for the final evaluation. In DEX, the aggregation procedure is defined by *decision rules,* an example of which is shown in Fig. 3.

	Ease of use	Communication	Functional environment	Help	Student's learning environment
	39%	29%	21%	11%	
1	low	low	<=average	*	low
2	low	low	*	low	low
3	low	<=average	low	*	low
4	low	<=average	<=average	low	low
5	<=average	low	low	low	low
6	>=average	high	high	high	high
7	high	>=average	>=average	high	high
8	high	>=average	high	*	high
9	high	high	*	high	high
10	high	high	>=average	*	high

Fig.3: Utility function for criterion, Student's learning environment

Obviously, there are many more such rules in the model. For each aggregate criterion (such as *Student's learning environment),* a similar table is defined. In the entire model there are 108 rules defined in this way.

- *Description of Variants*

The multi-attribute decision making model was tested on three Learning management systems: *Blackboard 6* (www.blackboard.com), *CLIX 5.0* (www.im-c.de) and *Moodle 1.5.2* (www.moodle.org). Blackboard is among the most perfected and complex LMSs on the market. The system offers various communication options (both synchronous and asynchronous) within the learning environment. The Blackboard LMS is designed for institutions dedicated to teaching and learning. Blackboard technology and resources power the online, web-enhanced, and hybrid education programs at more than 2000 academic institutions (research university, community college, high school, virtual MBA programs etc. CLIX is targeted most of all at big corporations, because it provides efficient, manageable, connected and expandable internet-based learning solutions. This scaleable, multilingual and customizable software aims at providing process excellence for educational institutions. For educational administrators, CLIX offers powerful features for course management and distribution. Additionally, it provides personalized learning paths for students, a tutoring centre for lectures and a whole bunch of innovative collaboration tools for both user groups, e.g. a virtual classroom. Altogether, CLIX makes planning, organizing, distributing, tracking and analyzing of learning and teaching a smooth and efficient process. Moodle is a free, open source PHP application for producing internet-based educational courses and web sites on any major platform (Linux, UNIX, Windows and Mac OS X). The fact that it is free of charge is especially attractive for schools and companies which always lack resources for the introduction of new learning technologies. Furthermore, the Moodle system is not only price-efficient – it can easily be compared to costly commercial solutions on all aspects. Courses are easily built up using modules such as forums, chats, journals, quizzes, surveys, assignments, workshops, resources, choices and more. Moodle supports localization, and has so far been translated into 34 languages. Moodle has been designed to support modern pedagogies based on social constructionism, and focuses on providing an environment to support collaboration, connected knowing and a meaningful exchange of ideas. The greatest disadvantage of the system is certainly support to e-learning standards, which is reflected on Fig. 4, showing evaluation results according to different assessment criteria for Blackboard, CLIX and Moodle Learning management systems.

3.3 LMS Evaluation and Analysis

The evaluation is carried out according to the tree of criteria from the basic criteria up. The method of aggregation is determined by the decision rules. The variant which is awarded the highest grade should be the best one.

Due to the complexity of LMSs and a large number of criteria it is essential that the decision-making model allows us to obtain not only the final assessment, but also a detailed partial analysis of individual elements. In this way we can detect weak points and disadvantages of the system, which can be used as the basis for system improvements. We can anticipate how specific criteria improvements would influence quality and we can achieve a more optimal distribution of resources at our disposal. The immense importance of individual criteria and their autonomy prevents the average of one or more criteria to automatically become the average of the entire system. For example, an LMS that was awarded an average grade in all three criteria (e.g. average *Student's learning environment*, *System, technology & standards, Tutoring & didactics*), cannot be called average, because it could be even worse, under average. On the other hand, a system with excellent technological and standardization solutions cannot be considered of high quality also from the methodology and didactics point of view, if the system does not provide an adequate *Student's learning environment*, which is essential for e-learning users, since it does not fulfill their objectives. Besides these, there are also some excluding factors that must be met in order for an

LMS to achieve a certain level from the point of view of security and privacy for example. We can renounce the safe SSL transfer in order to enhance the operational speed (this is especially important for users still using modems to connect to the Internet) and consequently positively influence applicability of the system. However, such a system does not meet security requirements, which are important in e-learning (they are considered important also by the decision-making model). The advantages and disadvantage of the systems are reflected in Fig. 4, showing evaluation results according to attributes: *Functional environment*, *Ease of use*, *Course analysis*, *Tutoring & didactics*, *Assessment* and *Standards support* for Blackboard, CLIX and Moodle LMS.

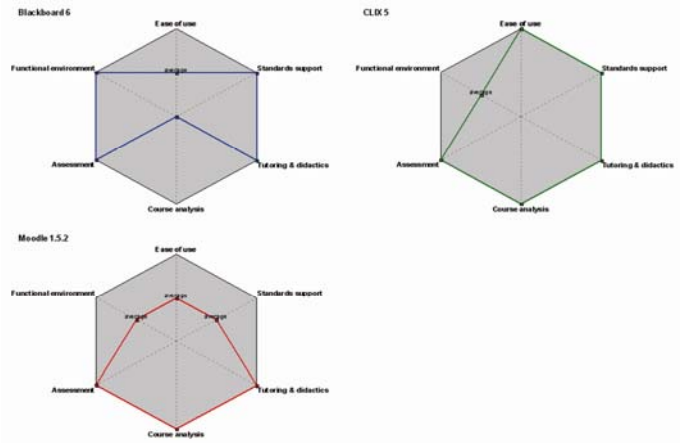

Fig. 4: Evaluation results for Blackboard 6, CLIX 5.0 and Moodle 1.5.2.

3.4 Usability Evaluation of CLIX 5.0

According to the results of the first phase of ECM (criteria evaluation), where the LMS CLIX 5.0 got the highest assessment among three different Learning Management Systems and methodology of Evaluation Cycle Management, we performed the second phase of evaluation: Usability testing (usability evaluation) of the LMS CLIX 5.0.

Standard user test procedures were adopted and were conducted by the respective evaluation administrators, who were responsible for recording the data, transcribing think-aloud protocols of the participants, asking them to fill out pre-test, post-task and post-test questionnaires and participant debriefing [11, 12, 17].

- *Participants*

We selected 10 participants, who were typical representatives of the target audience or were matched as closely to the criterion as possible. They possessed a certain level of experience and knowledge of information and communication technologies (ICT), experience in using software applications and had some basic knowledge about e-learning.

- *Test Objectives*

Usability goals for the LMS CLIX 5.0 were referred to on the hypothesis that users can utilize the services of the observed learning management system quickly, easily and accurately to accomplish their tasks in the way to attain the optimum level of effectiveness and efficiency, and find the navigation design comprehensive and user-friendly to attain an optimum level of satisfaction.

- *Task Scenarios*

A set of seven tasks covering the core functionalities of the LMS CLIX 5.0 and also presenting the potential usability problems. Here below is the list of the tasks:
(T1) – Updating a personal profile
(T2) – Booking a course
(T3) – Accessing the learning resources of a booked course
(T4) – Taking an e-test
(T5) – Joining and contributing to the discussion forum
(T6) – Making a post in the Chat-room
(T7) – Creating a personal Bookshelf

Each of the above seven tasks was translated into task scenarios, which render the test more realistic and problem oriented.

- *Measurements and Usability Metrics*

Quantitative Data

a) *Effectiveness*: completion rate (percentage of participants who completed each task correctly with/without assistance from a usability administrator), errors (number of errors: in menu choice, selecting an item from a list and other) and assists (number of times looking up on-line help and from a usability administrator).
b) *Efficiency*: task time (mean time of completion of each task, range and standard deviation of times) and completion rate efficiency (mean task time only for unassisted tasks).

Qualitative Data

c) Satisfaction: ratings and comments obtained through ASQ, CSUQ questionnaires and participant debriefing.

The results of the Usability Test

Quantitative Data

a) Effectiveness and Efficiency
Each participant was required to perform 7 tasks and fill out enclosed questionnaires. The mean time over 7

tasks is 47.65 minutes, with the range from 40.74 to 55.02 and a standard deviation of 4.453. Altogether, participants performed 70 tasks, 59 (84,3%) were correctly completed without assistance and 11 (15,7%) with assistance (on-line help, advice) from the usability administrator. Task 4 (Taking an e-exam) was found to be the most problematic. The mean time for Task 4 is 11.27 minutes and exceeds the time for completing the task (9 minutes), assessed by the experienced user, by 25,2%.

Furthermore, effectiveness and efficiency per task were computed. Effectiveness presents the rate that a task is correctly completed without intervention from the usability administrator or any other help. Efficiency is presented as the percentage of task completion per minute and it is calculated through dividing an unassisted completion rate by the corresponding unassisted average time-on-task. The average effectiveness (Table 4) over 7 tasks is 84,28%, ranging from 70% (Task 4, 5, 7) to 100% (Task 2, 3). The average efficiency over 7 tasks is 14.94%/minute, ranging from 6.78%/min (Task 4) to 25.32%/min (Task 3). From these data it is evident once again that Task 4 potentially should be problematic.

b) Errors
There are three types of errors that the participants may commit: Menu choice error (M), Select from list error (L) and Other error (O). As shown in Table 5, the total number of errors committed by 10 participants is 90, with the Menu choice Error being the highest (53), followed by Select from list error (30) and Other error (7). The average number of errors over 10 participants is 9, ranging from 7 to 16. One of the measured attributes was also a frustration per task and Task 4 has caused much higher frequency of frustration (4) than the other tasks. These results are consistent with other findings that Task 4 evoked the highest number of errors and amount of frustration and had the longest time-on-task.

Qualitative Data

a) After Scenario Questionnaire (ASQ)
The ASQ was developed to be used immediately following scenario completion in scenario-based usability studies. The three questions of ASQ unequivocally measure one single underlying aspect of participants' perceptions of how easily and quickly the scenarios were completed and the contribution of support information to carrying out the tasks. Each item is rated with a 7-point Likert scale, with 1 being "Strongly agree" and 7 "Strongly disagree". The items are phrased in a positive manner. The overall satisfaction rate over 10 participants is above the average with the mean value of 2.14. The lowest rating of satisfaction, with the mean value 4.67, was for Task 4, which indicates that this functionality is difficult and complex for general users to manage.

b) Computer System Usability Questionnaire (CSUQ)
CSUQ is publicly available questionnaire which contains 19 questions with a 7-point Likert scale for each answer. CSUQ can gauge three factors of satisfaction: *System Usefulness* (SYSUSE), *Information Quality* (INFOQUAL), and *Interface Quality* (INTERQUAL). A higher score gained out of a 7-point Likert scale means higher satisfaction with the system. The average overall satisfaction over 10 participants is 5.68, with a standard deviation 0.48. The implication of the value is that the users' general satisfaction with the LMS CLIX 5.0 is good. Both the system usefulness and interface quality were above the system average, whereas the information quality was a little bit below the average. A little lower assessment for the INFOQUAL should be attributed to the fact that most of the 10 participants were not satisfied with the supporting on-line help and feedback massages [10, 11, 15, 16].

c) Participant debriefing
The debriefing session is an extremely important portion of the usability evaluation as it allows participants to convey their exact feelings about the product being tested. In addition, it allows the usability administrators to ask direct questions about murky points of each participant's evaluation, i.e. any sticking points they encountered during the testing, any specific problems with wording, etc. From these interviews usability administrators were identified some usability problems and acquired very useful information and recommendation from the real users (e.g. not enough clear taxonomy, lack of on-line help, the graphical user interface is confusing due to many colors and fonts etc.).

4 Analysis of Results and Findings

The main goal of the case study was the selection of the most suitable and appropriate LMS among the three available (BlackBoard 6, Moodle 1.5.2 and CLIX 5.0), which would to the greatest degree possible, satisfy the requirements and needs of the target group.

As was already expected at the commencement of evaluation, a system which would entirely satisfy the target group of users was extremely difficult to find.

Each system observed had its strengths and weaknesses, thus the choice of the most suitable system was that much harder. The ECM methodology in the first phase – development of a multi-attribute decision-making model – foresees the choice of only one of a number of solutions, namely that which best satisfies the criteria defined especially for the aforementioned target group. Furthermore this solution, selected as most suitable in the second ECM phase was then also validated by testing its usability on end users whereby the actual usefulness of the system was ascertained confirming that the LMS chosen offers sufficient and proper support for execution of the exercises which will lead to the realization of the planned objective. The advantage of ECM lies in the fact that only one – the most suitable – solution (LMS) is chosen during the first phase of evaluation allowing a decrease in costs and time used, for continued evaluation of its usability is subject to only one solution and not all.

Based on the results acquired with the aid of the first phase ECM methodology (criteria evaluation) it is evident that the LMS CLIX 5.0 obtained the best marks of all three main criteria, at the same time coming closest to the criteria of an optimal solution. Since it was, however, the second ECM phase (usability evaluation) which supplied the answer of whether the selected LMS CLIX 5.0 was really the most suitable solution for the selected target group, it was additionally subjected to the testing of its usability. Ten participants participated in the test, which, on the basis of 7 tasks, verified the key functionalities of the system. The LMS CLIX 5.0 also proved to be an extremely suitable system for the target group of users in the second phase of evaluation according to ECM methodology. While performing usability testing several deficiencies were ascertained which, according to experts, represent merely minor corrections *(e.g. facilitation of navigation to e-testing, improvement of on-line help features, facilitation of terminological support texts in on-line documents, better color reconciliation and fonts for the user server, etc.*).

Based on the results of the study implemented according to ECM methodology we concluded that the selected LMS, CLIX 5.0 (with several minor corrections) was suitable for the chosen target group of users – employees in small and middle-sized companies and was also recommended for use by us. Since CLIX 5.0 received good marks both in the first and second phases of ECM and in terms of feedback, consequently modification of the multi-criteria decision-making model was not required.

References:

[1] Axup Jeff: *Comparison of Usability Evaluation Methods*.[URL:http://www.userdesign.com/usability_uem.html], 3.6.2007

[2] Baker, D., Bridges, D., Hunter, R., Johnson, G., Krupa, J., Murphy, J. and Sorenson, K.. *Guidebook to Decision-Making Methods*, WSRC-IM-2002-00002, Department of Energy, 2002.

[3] Bohanec, M. & Rajkovič, V.. *DEX: an expert system shell for decision support*, Sistemica 1, 1990, 145–157.

[4] Bohanec, M., Rajkovič, V.. *Multi-Attribute Decision Modeling: Industrial Applications of DEX*, Informatica, 23, 1999, 487−491.

[5] Chankong, V. & Haimes, Y.Y.. *Multi-objective Decision Making*: Theory and Methodology. North-Holland, 1983.

[6] Debevc Matjaž, Kocjan Stjepanovič Tanja: *Uvod v oblikovanje interakcije človek-računalnik*, Maribor: Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru, 2005, pp. 316.

[7] Dumash, J., Redish, J.C.. *A Practical Guide to Usability Testing*. Intellect Books Ltd, 1999, pp. 97–263.

[8] Harker Susan: *Review, Report and Refine Usability Evaluation Methods: Standardisation in the field of Usability*. Athens: The 3rd COST294-MAUSE International Workshop, 2007, pp. 7-11.

[9] Hémard, D.. *Language Learning Online: Designing Towards User Acceptability*, in Felix, U. (ed.) Language Learning Online: Towards Best Practice, Lisse, 2003, pp. 21–42.

[10] Holzinger, A.. *Usability Engineering Methods for Software Developers*. Communications of the ACM, Vol. 48, 2005, pp. 71–74.

[11] Hom James: *The Usability Methods Toolbox*. [URL:http://jthom.best.vwh.net/usability/usahome.htm], 18.11.2006

[12] Humphreys, C.P. & Wisudha, D.A.. *Methods and Tools for Structuring and Analysing Decision Problems, Decision Analysis Unit*, The London School of Economics and Political Sciences, Tech. Rep. 87-1, London, 1987.

[13] Kirkpatrick, D.. *Evaluating Training Programs*. San Francisco, CA: Berrett Koehler Publishers, Inc.,1994.

[14] Law, E.. *International Usability Tests on the Multilingual EducaNext Portal.* Universal Exchange for Pan-European Higher Education, Usability Test Report, v. 2. 0., 2003.

[15] Lewis, J. R.. *IBM Computer Usability Satisfaction Questionnaires: Psychometric evaluation and instructions for use*. International

Journal of Human-Computer Interaction, 7(1), 1996, 57–78.

[16] Nielsen, J.. *Usability Engineering*. Cambridge, MA:Academic. 1993.

[17] Usability First: *Introduction to Usability*. [URL:http://www.usabilityfirst.com/intro/index.txl], 27.11.2006