Usability Evaluation Methodology for Web-based Educational Systems

Andrina Granić, *Vlado Glavinić** and *Slavomir Stankov*

Faculty of Natural Sciences, Mathematics and Education, University of Split
Nikole Tesle 12, 21000 Split, Croatia
Phone: (385) 21 38 51 33/134, Fax: (385) 21 38 54 31
E-mail: {*andrina.granic*|*slavomir.stankov*}*@pmfst.hr*

* Faculty of Electrical Engineering and Computing, University of Zagreb
Unska 3, 10000 Zagreb, Croatia
Phone: (385) 1 612 9955, Fax: (385) 1 612 9653, E-mail: *vlado.glavinic@fer.hr*

Abstract. Over the past decade the field of education has experienced the introduction of the Internet, which has radically altered the way humans teach and learn. In conjunction with other advances within ICT, here including multimedia, networking and software engineering, this has enabled the appearance of new generations of computer-based educational systems. Inclusion of intelligence and adaptivity has additionally led to the development of Web-oriented educational systems like Web-based intelligent authoring shells, which provide the means for automated generation of emulators of human teachers in the process of learning and teaching. Our research is focused on user interfaces for Web-based intelligent authoring shells, which enhance with adaptive and "Web-based" features the operating capabilities of legacy "on-site" systems. In the paper we describe the methodology for such Web-based shell's usability evaluation, along with results thus achieved. The methodology itself is composed of three usability evaluation methods including a scenario-based end-user testing, guidelines set for system review and a user-interaction satisfaction questionnaire.

1. INTRODUCTION

A novel approach to education based upon electronic technology, usually called e-learning, comprises different ways of computer support where teaching material can be delivered synchronously (e.g. Web-based videoconferencing, audio conferencing with presentation material, on-line chat) or asynchronously (e.g. computer managed instruction, intelligent tutoring systems, learning management instruction, learning content management systems). Learning Management Systems, LMSs, and Learning Content Management Systems, LCMSs, make nowadays the central point of interest in asynchronous delivery of teaching materials. The primary goal of LMSs is learner management, keeping track of their progress and performance across all activities in the learning and teaching process. LCMSs capabilities include management of either content or learning object, which is provided to the right learner at the right time. Although both LMS and LCMS, along with their derivative iLMS [Yacef 03], are all intensively investigated, it seems that Web-based intelligent authoring shells are better conceived and already offer all what is intended to be achieved with iLMSs [Brusilovsky 03].

We consider that Intelligent Tutoring Systems, ITSs, still represent the best way to enable one-to-one instruction [Fletcher 03] and at the same time is the best attempt to solving the 2-sigma problem, as pointed out by Bloom in his comparison of traditional classroom teaching vs. the individual one [Bloom 84]. It has been claimed that from the 70's, when the Scholar system was developed [Carbonel 70], ITSs systems undoubtedly have improved the process of learning and teaching for arbitrary domain knowledge [Fletcher 03], also taking into

account student individuality. While interacting with an ITS in a relaxed manner a student gets her/his own "computer teacher", which in comparison to a human tutor has no emotions or feelings.

This attitude is supported by our previous experience in researching ITSs and authoring shells, ASs, the latter providing automated ITS generation. Within these activities we conceptualized, developed and deployed a series of educational systems like *Tutor-Expert System*, TEx-Sys [Stankov 97], *Distributed Tutor-Expert System*, DTEx-Sys [Rosić 00] and *Adaptive Knowledge Base Builder*, AKBB [Granić 02]. The next member of the TEx-Sys family is *eXtended Tutor-Expert System*, *xTEx-Sys*, which is a Web-based intelligent authoring shell, presently under development.

It is obvious that remote access to ITSs represents a starting point for the development of e-learning, as one of the most important services an information infrastructure has to provide [NII 94]. In order to ensure a usable e-learning service we study design of user interfaces for Web-based intelligent tutoring systems, and specifically consider Web-based authoring shells accessible through standard Web browsers. This of course enables the well-known client-server paradigm, whose graphical user interface shares both similarities and differences with educational systems.

Within this framework *xTEx-Sys*, which will offer adaptivity and intelligence, will be developed according to usability engineering principles for Web-based systems. Consequently, a proper methodology is required, by means of which the relevant HCI attributes are quantified thus enabling system validation and especially identification of inherent weaknesses. In the following we discuss Web-based usability issues, and in particular the relevant ones concerning usability evaluation of our current shell's Web-based design. We also corroborate the evaluation methodology developed for this purpose along with the obtained results. The methodology applied is composed of a scenario-based usability test, a guideline evaluation and a usability questionnaire.

2. USER INTERFACES FOR WEB-BASED EDUCATIONAL SYSTEMS

Over the past decade the field of education has witnessed the introduction of Internet, a new and revolutionary technology that seems to radically alter the way humans teach and learn. Consequently, Web-based education has become nowadays a hot research and development area [Brusilovsky 98] with a challenging goal in the development of adaptive Web-based educational systems. This goes along with simultaneous efforts in including an increasing level of intelligence in computer-supported learning and teaching systems, what has led to the development of *Intelligent Tutoring Systems*, ITSs, whose principal operating paradigm is the imitation of human tutor capabilities [Fleischman 00]. ITSs thus take into consideration the knowledge about *what to teach* (domain knowledge), the *way to teach* (teacher knowledge), as well as the relevant information about the *student being taught* (student knowledge).

In order to further ease and automate the preparation of specialized ITSs, and hence to cover a variety of different domains of interest, ITS generators were developed, which are usually denoted as *Authoring Shells*, ASs [Murray 99]. ASs are intended to adjust to teachers as well as to students within an interactive learning environment by supporting teachers in the development of a series of ITSs for arbitrary domain knowledge and conversely, by enabling students to learn, test themselves, and be advised on further work.

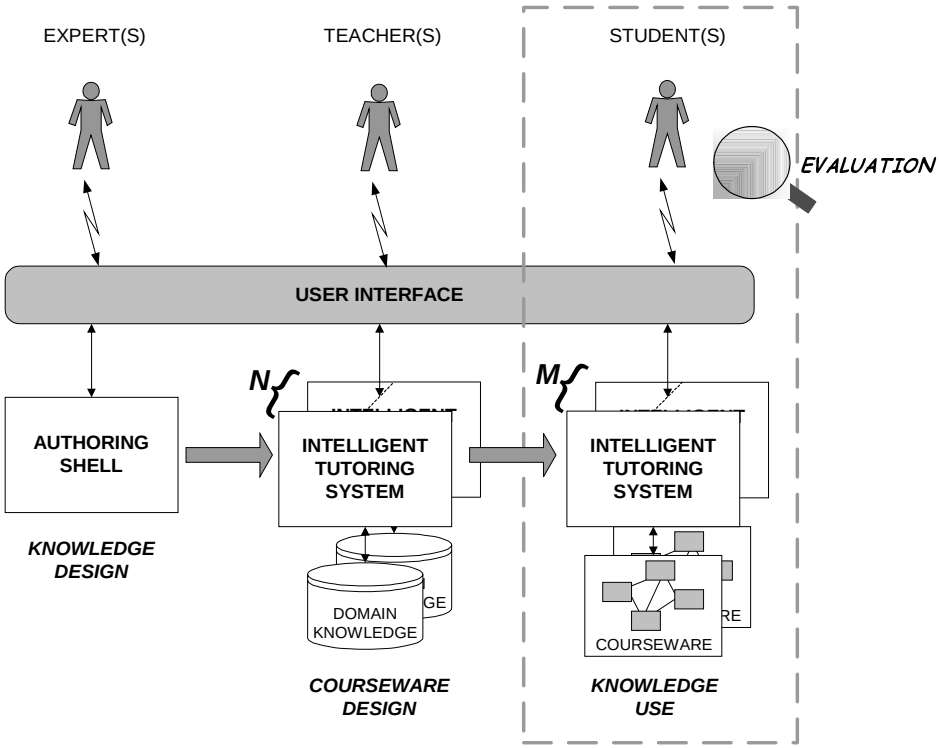

Figure 1: Relationship of roles in knowledge management for intelligent learning and teaching

Figure 1 illustrates the relationship among various actors in the process of learning and teaching. The process consists of three phases, which are sequenced, and represent domain knowledge specialization. Firstly, (one or more) experts for some particular domain knowledge by using an authoring shell create a domain knowledge base for an intelligent tutoring system. Of course, various experts can generate a number of such systems, each of them covering a specific knowledge domain, hence we could say that ITSs can be built by parameterizing some AS with specific domain knowledge bases, what is the *knowledge design phase*. The second phase involves a teacher who selects the appropriate knowledge for his course(s), here including a particular course structure, hence creating the respective courseware. This phase is thus denoted the *courseware design phase*. Finally the student(s), who is (are) courseware (i.e. knowledge) "users", "consume" this knowledge by apprehending it. Thus this phase is denoted *knowledge use phase*. As the investigated system *TEx-Sys* is an authoring shell it comprehends the functionality of all the three phases and its user interface exports to the users an adequate "look and feel" for all the three actor classes. In fact this user interface is usually considered to build an interaction shell of which in this paper we present only the characteristics of the interaction specific for the latter phase of the learning and teaching process.

3. USABILITY EVALUATION OF WEB-BASED INTERACTIVE SYSTEMS

Methodologies for building usable and functional computer systems have been introduced and refined over the past twenty years within the discipline of HCI. In order to achieve a system's almost transparency and enable users to fully concentrate on the work, HCI

principles include an early and consistent focus on end users and their tasks, empirical measurements of system usage, as well as iterative development. Usability, as one of the key system features, is primarily concerned with making a system easy to learn and easy to use. Studies show that redesigning a system's user interface on the basis of user testing (i.e. interaction measurement between users and computer systems) and iterating can substantially improve usability, because usability can only be meaningfully measured during task execution. Hence, the most promising approach to the generation of usable systems is to iterate design and usability evaluation until a satisfactory solution is achieved [Nielsen 93], [Shackel 91].

Although usability engineering has come to play an increasingly important role in conventional interactive system development, it is rarely part of Web-based system development [Levi and Conrad 97], despite the fact that there already exist Web style design guidelines (cf. [Tognazzini 03], [Levi and Conrad 98], [Borges *et al.* 96]). On the other hand, employing usability guidelines by themselves does not guarantee the development of usable systems, consequently usability evaluation, as a distinct validation process, must be performed. In the case of Web-based systems this means to take into consideration similarities and divergences with respect to conventional systems. Many different usability evaluation methods can be employed to elicit information on Web design aspects [Levi and Conrad 98], [Tullis 98], [Keevil 98], [Morkes and Nielsen 98], [Borges *et al.* 96], [Nielsen 93]:

- heuristics, i.e. design principles, can be used by experts to judge usability,
- benchmarking can be used to compare one Web site with another or against a set of standards,
- prototyping can be used to quickly and cheaply develop a mock site that can be shown to users before the real site is launched,
- a Web site can be evaluated against a checklist of usability items, or
- users can participate in focus groups or in controlled laboratory sessions in order to provide feedback on the usability of the site.

However, it is important to point out that any kind of usability evaluation will improve the final version of the system, as long as its results provide an appropriate feedback on which further improvements could be achieved.

4. METHODOLOGY FOR WEB-BASED INTELLIGENT AUTHORING SHELL EVALUATION

Within the research on user interfaces for Web-based intelligent ASs, we consider the development of a suitable usability evaluation methodology that should enable the quantification of relevant HCI attributes and in turn AS validation and especially identification of inherent weaknesses.

This methodology is of course derived from the one used to conduct usability evaluation of the "on-site" *TEx-Sys* authoring shell [Granić and Glavinić 02], which consisted of user testing during users' walkthrough along the AS interface, guided by a set of predefined steps. We chose such an approach having in mind that usability could only be meaningfully measured during task performance and that it would be better to perform any kind of usability measurement than no testing at all. The methodology is based on criteria expressed in terms of objective performance measures in systems use, as well as in terms of users' subjective assessment.

Since designing Web-based systems is different than designing traditional interfaces, an appropriate evaluation methodology is required, which will enable system validation on the one hand and provide means to compare achieved results with those obtained by evaluating previous system versions within the evolutionary design cycle. We start with the evaluation of the current AS distributed version *DTEx-Sys* (see Figure 2 for a snapshot of the system), which was developed without regard to HCI principles of usable design and without applying usability evaluation. This same usability evaluation methodology will afterwards be applied in evaluating of the redesigned AS version *xTEx-Sys*.

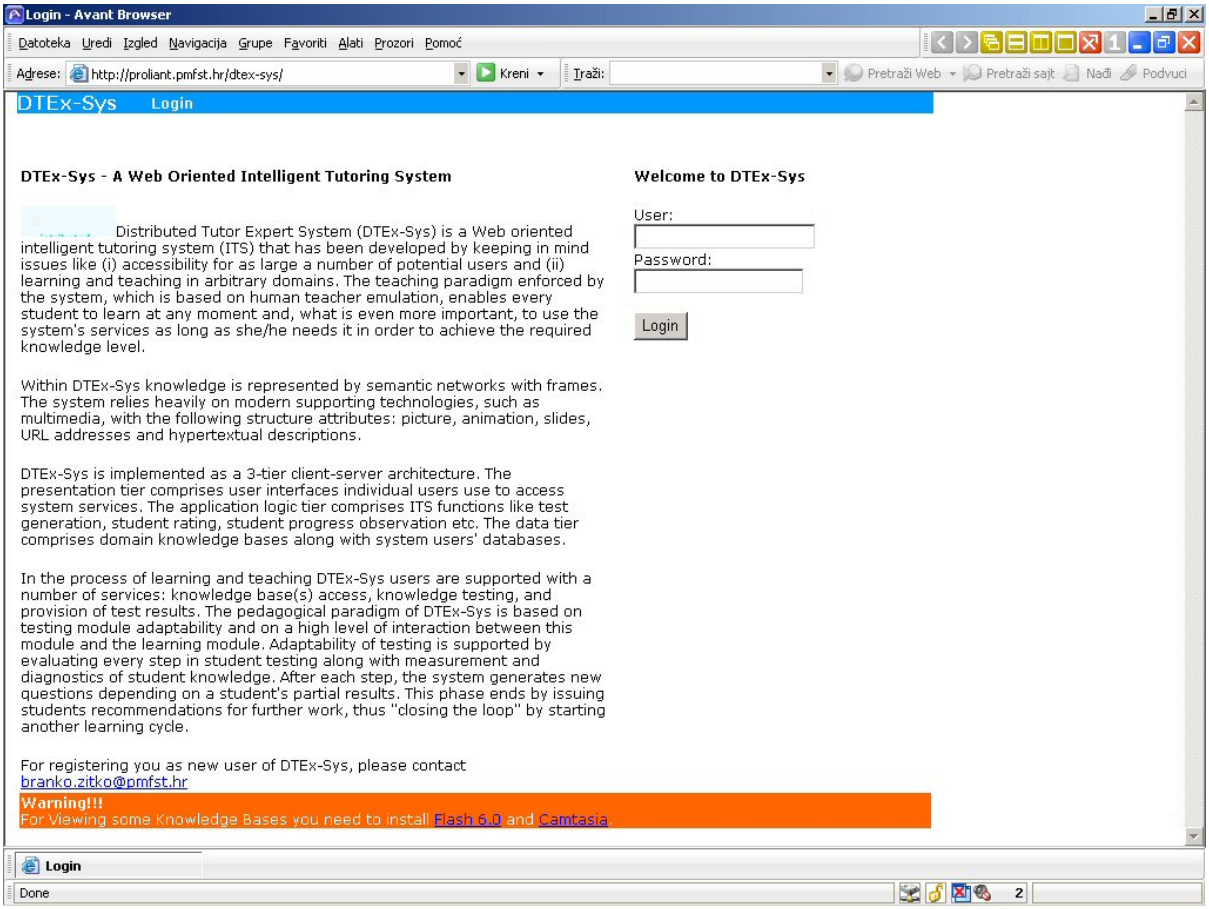

Figure 2: *DTEx-Sys* home page

4.1. Objective

In order to cope with the above requirements, the methodology for Web-based intelligent AS usability evaluation must obviously be composed of several usability evaluation methods, covering specific objectives:

- to determine the usability shortcomings of the current distributed shell's version, and also to enable the comparison of results obtained from testing several user categories,
- to determine improvements hopefully achieved by the new *xTEx-Sys* shell,
- to determine whether this, or some slightly modified methodology, is a promising one for the evaluation of other (types of) Web-based educational systems.

4.2. Methodology

Our methodology is composed of three usability evaluation methods: (i) a scenario-based usability test, (ii) a guideline evaluation and (iii) a usability questionnaire, providing means:

- to perform a scenario-based end user testing,
- to review the system using a set of metrics/guidelines and
- to probe the user subjective satisfaction with the interaction.

Empirical testing of end users with a *scenario-based usability test* enables testing site usage, as well as page design. Generally, the results of scenario-based usability tests are tabulated using such measures as whether the participants have correctly accomplished the tasks, the time taken for each task or/and number of pages accessed for each task. A *set of guidelines*, used as a checklist, also enables testing page design and site usage because users are in the position to identify usability problems and classify each problem found as a violation of one or more usability principles while performing given tasks or scenarios. In addition to evaluating hard measures by means of scenario-based testing, it is extremely useful to investigate those aspects of interface design that contribute to users' subjective feelings of satisfaction or frustration through a *usability questionnaire*.

A major "strength" in this approach is achieving results both from guideline evaluation and from empirical user-based evaluation, enhanced with users' feedback on their comfort while working with the shell. This reasoning is corroborated with literature claims that heuristic evaluations detect between 40% and 60% of the usability problems an empirical user test would find, and that the types of problems found are roughly comparable, e.g. [Nielsen and Molich 90].

4.2.1 A scenario-based end user testing

A *scenario-based usability test* involves representative AS's end-users and scenarios or specific tasks designed to cover the major shell's functionality and to simulate expected real-life usage patterns. Although measures such as task correctly accomplished or/and task completion time are usually achieved, it is much better to perform some more crucial measurements like how much users actually learn of the information (in fact domain knowledge) contained in the authoring shell's site. Consequently, a scenario-based end user testing enables us to measure:

- correctness/accuracy of tasks,
- recognition and recall memory,
- how much and in what time users actually learn.

The third aspect is very important because we are dealing with educational systems intended to improve the process of learning and teaching by attempting to mimic the capabilities of human teachers. Namely, communication between users and ITSs/ASs is inherently complex, especially when supporting student interaction because of the students' dealing with concepts (domain knowledge) yet not understand well. Therefore it is obvious that shell's interface, as well as an interface of any other computer system, should be transparent and do not bother users. In order to obtain these measures, a scenario-based usability test comprises the following expected usage patterns:

- three search tasks,
- a short examination after some time spent on the site,
- an AS's quiz enabling to test the achieved knowledge.

The great advantage of such empirical end-user testing is that achieved results are unquestionable. Unlike heuristic evaluation, where HCI experts speculate as to what may cause the users difficulties, end-user testing will highlight the place where users actually do have difficulties. When performing a scenario-based usability testing we are taking into account the fact that the best results come from testing no more than 5 users, because they can find 85% of the usability problems [Nielsen 00].

4.2.2 A set of usability guidelines

Of the eight inspection methods, *heuristic evaluation* is the least formal, and involves having usability specialists judge whether aspects of a given interface conform to a list of established usability principles (heuristics). Although an exact number of evaluators for heuristic evaluation depend on a cost-benefit analysis, a usage of three to five usability specialists is recommended, because they can usually identify about 75% of the usability problems [Nielsen 03]. However, in order to overcome the problem of not having enough usability specialists that can be involved in usability evaluation, we decided to perform a "less formal" guideline evaluation, conducted by a group composed of five participants.

The AS distributed version is evaluated using our own set of metrics/guidelines derived through an analysis of the literature on Web-based computer systems evaluation, e.g. [Bevan 98], [Levi and Conrad 98], [Nielsen 96], [Jones and Okey 95], among which just one set of guidelines (the last one) dealing with interface design evaluation for computer-based educational systems. Because the majority of authors do not suggest assigning marks to the set of their guidelines, the assignment of scores is strictly arbitrary, but when coupled with scenario-based end user testing and questionnaire for user interaction satisfaction, it should nevertheless provide an accurate and objective usability assessment. The set of guidelines is applied as a checklist so that the participants had to respond whether the shell's site was to be considered to full-fill the guideline ('*Done*') or more work was needed ('*Has to be done*') on a seven-point scale. '*Done*' was scored as 7, so that the greater the average on the guideline scale, the better the evaluated aspect of Web-based authoring shell is. The list of guidelines is given in Figure 3.

1. Design an effective home page that will establish the site identity and give a clear overview of the content.
2. Structure information hierarchically so that it is meaningful to the user.
3. Use a consistent page layout and indicate similar concepts through identical terminology and graphics.
4. Integrate the information across different media types.
5. Use terminology familiar to the user.
6. Design for recognition rather than recall - make actions and options visible, do not rely on the user remembering information.
7. Make a pleasing and minimalist design - avoid long text and do not include irrelevant and distracting information.
8. Provide users with information to let them know where they are and where they can go.
9. Provide visual effects to give users visual feedback that their choices have been performed and registered by the program.
10. Provide links on each page to a list of local content, a site map and home.

Figure 3: Set of usability guidelines

4.2.3 A questionnaire for user-interaction satisfaction

In addition to evaluating "hard" measures like time to complete a task and error rates, it is extremely useful to investigate the less observable aspects of interface design that cumulatively contribute to a user's subjective feelings of satisfaction or frustration. Questions in the questionnaire for user satisfaction are formulated according to the ones from the literature (cf. *QUIS* [Harper and Norman 93], *IBM Computer Usability Satisfaction Questionnaires* [Shneiderman 98], *SUMI* [HFRG 93], WAMMI [WAMMI consortium 02]). Furthermore, two important facts were also taken into account – it is a design of a Web-based interface and, what is more important, it is an interface of a Web-based educational system that is under evaluation. The most intelligent system in the world does no good if users avoid it because they find it annoying. The *usability questionnaire* supports testing and determination of user subjective satisfaction with the shell's interface, as well as her/his satisfaction with its ease of use, efficiency, likebility, as well as with the attitude the system induces in users during its usage (see Figure 4). Participants indicate level of their agreement with a questionnaire statement on a seven-point Likert scale. From the standpoint of the

	For each word below, please indicate how well it describes the site:		
1	annoying	disagree • • • • • • •	agree
2	confusing	disagree • • • • • • •	agree
3	frustrating	disagree • • • • • • •	agree
4	interesting	disagree • • • • • • •	agree
5	stimulating	disagree • • • • • • •	agree
6	tiresome	disagree • • • • • • •	agree
7	useable	disagree • • • • • • •	agree
8	unpleasant	disagree • • • • • • •	agree
9	I feel in control when I am using this site.	disagree • • • • • • •	agree
10	This site uses terms understandable and familiar to me.	disagree • • • • • • •	agree
11	This site needs more introductory explanations.	disagree • • • • • • •	agree
12	I find this site useful.	disagree • • • • • • •	agree
13	Everything on this site is easy to understand.	disagree • • • • • • •	agree
14	This site is too slow.	disagree • • • • • • •	agree
15	I get what I expect when I click on objects on the site.	disagree • • • • • • •	agree
16	It is difficult to move around this site.	disagree • • • • • • •	agree
17	I feel efficient when using this site.	disagree • • • • • • •	agree
18	Compared to what I expected, the tasks did go really quickly.	disagree • • • • • • •	agree
19	I will characterize this site as an innovative one.	disagree • • • • • • •	agree
20	Overall, I am quite satisfied with this site.	disagree • • • • • • •	agree

Figure 4: Questionnaire for user-interaction satisfaction

single participant the responses represent her/his subjective opinion, but as an average value taken from all participants they indicate an objective value of Web-based AS pleasantness.

5. EVALUATION AND RESULTS

5.1. Participants

The evaluation was carried out by one experimenter. In order to get ready for the evaluation session, the experimenter prepared several documents: (i) an evaluation process overview describing the objectives and target audience, (ii) expected Web-based AS usage patterns/tasks along with search tasks and a short exam, (iii) a list of usability principles/guidelines and (iv) a questionnaire for user-interaction satisfaction.

The next step was the identification of participants evolved in an evaluation process. Nielsen recommends using three to five usability experts as evaluators when carrying out heuristic evaluation, as well as a group of five test users when performing usability testing. However, we had to overcome the problem of not having enough usability specialists that can be involved in usability evaluation. To broaden the evaluator/test user perspectives, but also to help us evaluate a shell's interface due to the fact that is better to perform any kind of usability measurement than no testing at all, we decided to run the evaluation process in parallel with three separate groups. The first was a group of five system developers (evaluators-novices in the usability field) who were involved in the development of the current AS's versions and were contributors in the development of an internal teaching material as well. The second evaluation group consisted of five students of mathematics and computer science from the fourth academic year and the third one of five students of computer science and polytechnic who are all familiar with the computer-based learning and teaching paradigm.

5.2. Procedure

The evaluation process was almost identical for every group of participants. The experimenter met with each group for 45 minutes to explain the purpose of the sessions, present both the methodology overview and the set of guidelines. Throughout the evaluation session, the participants received both printed instructions and verbal instructions from the experimenter. At the end of this initial briefing, an evaluation session was performed with each group. The experimenter was present to assist with any difficulties and to answer questions as they possibly arose.

The evaluation sessions lasted about one and half hours each. The participant began at the Web-based AS's home page. The first three tasks included the search for specific facts located on separate pages in the site, without using a search tool or the "Find" command. Next, the participants were instructed to spend 30 minutes learning as much as possible from site pages related to specific domain knowledge, as preparation for a short examination, as well as for a final task – answering questions about the learned specific domain knowledge which was individually generated by shell's quiz. While performing specific tasks, participants were identifying potential usability problems and were at the same time in the position to "tie" each problem found to the specific guideline it violated. The set of guidelines was used as a checklist so that the participants had to respond whether the evaluated shell's site was considered to have full-fill the guideline ('*Done*') or not, as already described before. Finally, participants were asked to fill in the questionnaire for user-interaction satisfaction indicating the level of their agreement with a statement on a seven-point Likert scale.

5.3. Measures

An applied methodology composed of three usability evaluation methods (scenario-based usability test, guideline evaluation and a usability questionnaire) provided the means for the following measures:

- ***task correctness*** was a percentage score based on the number of correct answers users gave in the three search tasks; e.g. one of the questions was: *Send the following message to the person or group responsible for registering new users of this site: "<Your name> was visiting this site at <current time> on 5th of December". What is the e-mail address to which you sent the message?*
- ***memory*** comprised two measures from the short examination:
  - *recognition memory* as a percentage score based on the number of correct answers to multiple-choice questions; e.g. one of the questions was: "*The term* computer system *from the domain knowledge is described with an additional information expressed in terms of: a) text, b) graphics or c) animation*." and
  - *recall memory* as a percentage score based on the number of correctly recalled answers; e.g. one of the questions was: "*Do you remember any additional information used to better describe some terms from the specific domain knowledge?*"
- ***achieved knowledge*** was the score obtained after performing a quiz on specific domain knowledge generated for every individual participant by the Web-based AS itself; the score for each participant group was an average value of all respective items;
- ***quiz solving time*** was the time participants spent in quiz solving; the score for each participant group was the average of all measured items;
- ***subjective satisfaction*** was determined from participants' answers to a paper-and-pencil questionnaire; some questions asked specific aspects of working with the site/shell, while others asked for an assessment of how well certain adjectives described the site (anchored by arrange between "*disagree*" and "*agree*"); all questions used a seven-point Likert scale and some of them were inversely coded; for each question the items were averaged so that the possible range was from 1 to 7; the subjective satisfaction was the mean score of the following four aspects (each comprised five items from the questionnaire):
  - *ease of use* of the site/shell; an example of a questionnaire item: "*Everything on this site is easy to understand.*",
  - *efficiency* of the site/shell; an example of a questionnaire item: "*This site is too slow.*" (this item was inversely coded),
  - *likability* of the site/shell; an example of a questionnaire item: "*Overall, I am quite satisfied with the site.*", as well as
  - *user feelings* while working with the site/shell; an example of a questionnaire items: "*frustrating*" and "*confusing*".

5.4. Accomplished results

Averaged scores for a given set of guidelines used as a checklist are given in Table 1, while main measurements from scenario-based usability testing, as well as from questionnaire for user-interaction satisfaction are presented in Table 2.

Evaluation found relatively few differences in main measurements while task performing, as well as in averaged scores for a set of guidelines between the first group consisting of system developers and the second one consisting of students of mathematics and computer science from the fourth academic year. More significant difference was found in scores obtained from the third group consisting of students of computer science and polytechnics. Much of this difference can be attributed to the system developer's expertise on the one hand and to quite good generation of students of mathematics and computer science on the other.

Table 1: Averaged scores for a set of guidelines

guidelines	Group #1	Group #2	Group #3
no. 1	4,2	3,8	3,6
no. 2	5,0	6,2	3,8
no. 3	6,5	5,2	4,4
no. 4	5,2	5,4	3,8
no. 5	6,5	6,2	4,4
no. 6	5,0	4,6	3,6
no. 7	6,2	6,6	4,4
no. 8	5,0	5,6	3,0
no. 9	5,5	3,4	4,6
no. 10	4,0	4,6	3,8

In general, due to shell's development without an employment of any usability evaluation, it is not surprising that every group of participants identified specific problems at all levels of the current distributed AS version. While system developers found that more has to be done on the provision of links on each page to a list of local content, a site map and home page (guideline no. 10), students of mathematics and computer science found that more has to be done on the provision of visual effects for giving users visual feedback (guideline no. 9). In addition, every group identified problems with the design of an effective home page that will

Table 2: Mean scores for five major measures

	Group #1	Group #2	Group #3
task correctness	91%	86,4%	79,6%
memory	80%	80%	55%
achieved knowledge	55/58	55,4/58	50,4/58
quiz solving time	3,75 min	6 min	6,4 min
subjective satisfaction	5,14	5,55	4,17

establish site identity (guideline no. 1), as well as with designing for recognition rather than recall (guideline no. 6). The violation of the latest guideline can also be recognized in the score for one of the five major measures – *memory* – because it is obvious that design must not rely on user remembering information.

As seen from accomplished measures, we did not measure *time to complete a task*, because we conceive that is better to perform more crucial measurements like how much users actually learn of the information contained in the shell's site – *achieved knowledge* – as well as do users feel efficient while working with the shell – *subjective satisfaction/efficiency* – (information obtained from questionnaire for user-interaction satisfaction).

6. CONCLUSION

Advances in information and communication technology, specifically in multimedia, networking and software engineering have enabled the apparition of a new generation of computer-based educational systems. Namely, over the past decade the field of education has witnessed the introduction of Internet, a new and revolutionary technology that seems to radically alter the way humans teach and learn. Inclusion of intelligence and adaptivity in computer-based learning and teaching systems, along with the employment of Internet and its technological capabilities, has led to the development of Web-oriented educational systems like Web-based intelligent authoring shells, which principal operating paradigm is the imitation of human tutor capabilities.

It is obvious that remote access to authoring shells represents a starting point for the development of e-learning, as one of the most important services an information infrastructure has to provide. In order to ensure a usable e-learning service we study design of user interfaces for Web-based intelligent authoring shells accessible through standard Web browsers. Within this framework *eXtended Tutor-Expert System – xTEx-Sys*, which will offer adaptivity and intelligence, is presently under development according to usability engineering principles for Web-based systems. Consequently, a proper methodology is required, by means of which the relevant human-computer interaction characteristics are quantified thus enabling system validation and identification of inherent weaknesses.

In the paper we discuss Web-based usability issues, and in particular the relevant ones concerning usability evaluation of our current shell's Web-based design. We also corroborate the evaluation methodology developed for this purpose along with the obtained results. The methodology applied is composed of a scenario-based usability test, a guideline evaluation and a usability questionnaire. Our experience indicates that useful usability validation with significant identification of inherent interface weaknesses can be performed quite easily and quickly, with relatively no cost except the employees' time. Ease of learning, ease of use, as well as general user satisfaction, along with quality and comprehensiveness of content and functional capabilities, will determine the success or failure of this approach.

ACKNOWLEDGEMENTS

This work has been carried out within projects 0036033 *Semantic Web as Information Infrastructure Enabler* and TP-02/0177-01 *Web-based Intelligent Authoring Shell*, both funded by the Ministry of Science and Technology of the Republic of Croatia.

REFERENCES

[Bloom 84] B. Bloom, The Two-Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring, Educational Researcher, 13, 1984, pp. 4-16.

[Brusilovsky 03] P. Brusilovsky, A Distributed Architecture for Adaptive and Intelligent Learning Management Systems, *AIED 2003 Workshop Proceedings Towards Intelligent Learning Management Systems*, 20th July 2003.

[Brusilovsky 98] P. Brusilovsky, Adaptive educational systems on the World Wide Web: a review of available technologies, in G. Ayala (ed), *Proceedings of Workshop "Current Trends and Applications of Artificial Intelligence in Education"*, at the 4th World Congress on Expert Systems, Mexico City, Mexico, ITESM, 1998, pp. 9-16.

[Bevan 98] N. Bevan, Usability Issues in Web Site Design, *Proceedings of the HCI International '97*, San Francisco, USA, 24-30 August 1997, Elsevier Press, 1997, also available as Version 3, 1998, `http://www.usability.serco.com/papers/usweb98.pdf`

[Borges *et al.* 96] J. Borges, I. Morales, N. Rodriguez, Guidelines for Designing Usable World Wide Web Pages, *Conference Companion of the CHI'96 Computer-Human Interaction Conference*, Vancouver, Canada: ACM, 1996, pp. 277-278.

[Carbonel 70] R. Carbonel, AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction, *IEEE Transactions on Man-Machine Systems*, MMS-11 (4), 1970, pp. 190-202.

[Fleischman 00] A. Fleischmann, The electronic teacher: the social impact of Intelligent Tutoring Systems in education, 2000, `http://wwwbroy.in.tum.de/~fleischa/papers/its.html`

[Fletcher 03] J. Fletcher, Evidence for Learning From Technology – Assisted Instruction, in H. O'Neil, R. Perez (eds), *Technology Applications in Education: A Learning View*, Lawrence Erlbaum Associates, 2003, pp. 79-99.

[Granić 02] A. Granić, *Foundation of Adaptive Interfaces for Computerized Educational Systems*. Ph.D. Diss., Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, 2002, (in Croatian)

[Granić and Glavinić 02] A. Granić and V. Glavinić, An approach to usability evaluation of an intelligent tutoring system, in N. Mastorakis, V. Kluev (eds), *Advances in Multimedia, Video and Signal Processing Systems*, Athens, Greece: WSEAS Press, 2002, pp. 77-82.

[Harper and Norman 93] B. Harper and K. Norman, Improving User Satisfaction: The Questionnaire for User Interaction Satisfaction Version 5.5, *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference*, Virginia Beach, VA, 1993, pp. 224-228.

[HFRG 93] HFRG, Human Factors Research Group, Software Usability Measurement Inventory, SUMI, 1993, `http://www.ucc.ie/hfrg/questionnaires/sumi`

[Jones and Okey 95] M. Jones and J. Okey, Interface Design for Computer-based Learning Environments, February 21, 1995, http://www.hbg.psu.edu/bsed/intro/docs/idguide/, [Also abridged version appears in B. Khan (ed), *Web-Based Instruction*, Educational Technology Publications, Englewood Cliffs, New Jersey, 1997]

[Keevil 98] B. Keevil, Measuring the usability index of your web site, in *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, Quebec Canada, 1998, pp. 271-277.

[Levi and Conrad 98] M. Levi and F. Conrad, Usability Testing of World Wide Web Sites, Research Papers of Bureau of Labor Statistics, 1998, http://stats.bls.gov/ore/htm_papers/st960150.htm

[Levi and Conrad 97] M. Levi and F. Conrad, Usability testing of World Wide Web sites: a CHI 97 Workshop, *SIGCHI Bulletin* Vol. 29, No. 4, October 1997, http://www.acm.org/sighchi/bulletin/1997.4/levi.html

[Morkes and Nielsen 98] J. Morkes and J. Nielsen, Applying Writing Guidelines to Web Pages, *Proceedings of ACM CHI'98 Conference on Human Factors in Computing Systems*, April 1998, Los Angeles, CA., 1998, pp. 321-322.

[Murray 99] T. Murray, Authoring Intelligent Tutoring Systems: An analysis of the state of the art, *International Journal of Artificial Intelligence in Education*, Vol. 10, 1999, pp 98-129.

[Nielsen 03] J. Nielsen, How to conduct a Heuristic Evaluation, 2003, http://www.useit.com/papers/heuristic/heuristic_evaluation.html

[Nielsen 00] J. Nielsen, Why You Only Need to Test With 5 Users, Alertbox March, 2000, http://www.useit.com/alertbox/20000319.html

[Nielsen 96] J. Nielsen, Top Ten Mistakes in Web Design, Jacob Nielsen's Alertbox for May 1996, http://www.useit.com/alertbox/9605.html

[Nielsen 93] J. Nielsen, *Usability Engineering*, Academic Press, Boston, MA, 1993.

[Nielsen and Molich 90] J. Nielsen and R. Molich, Heuristic evaluation of user interfaces, *Proceedings of ACM CHI'90 Conference*, Seattle, WA, 1-5 April, pp. 249-256. 1990.

[NII 94] NRENAISSANCE Committee: *Realizing the Information Future. Internet and Beyond*, National Academy Press, Washington, DC, 1994, http://www.nap.edu/readingroom/books/rtif/

[Rosić 00] M. Rosić, *Establishing of Distance Education Systems within the Information Infrastructure*, M.Sc. Thesis, Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, 2000 (in Croatian)

[Stankov 97] S. Stankov, *Isomorphic Model of the System as the Basis of Teaching Control Principles in the Intelligent Tutoring System*, Ph.D. Diss., Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, 1997 (in Croatian)

[Shackel 91] B. Shackel, Usability – context, framework, design and evaluation, in B. Shackel, S. Richardson (eds), *Human Factors for Informatics Usability*, Cambridge University Press, Cambridge, 1991, pp. 21-38.

[Shneiderman 98] B. Shneiderman, *Designing the User Interface. Strategies for Effective Human-Computer Interaction*, 3rd Edition, Addison-Wesley, Reading, MA, 1998.

[Tognazzini 03] B. Tognazzini, First principles, Nielsen Norman Group, 2003, http://www.asktog.com/basics/firstPrinciples.html

[Tullis 98] T. Tullis, A Method for Evaluating Web Page Design Concepts, *Proceedings of ACM CHI'98 Conference on Human Factors in Computing System*, April 1998, Los Angeles, CA, 1998, pp. 323-324.

[WAMMI consortium 02] WAMMI consortium, WAMMI – Web Usability Questionnaire, 2002, http://www.wammi.com

[Yacef 03] K. Yacef, Some thoughts on the synergetic effects of combining ITS and LMS technologies for the service of Education, *AIED 2003 Workshop Proceedings Towards Intelligent Learning Management Systems*, 20th July 2003.