A Comparative Study of Two Usability Evaluation Methods Using a Web-Based E-Learning Application

Samuel Ssemugabi
School of Information Technology
Walter Sisulu University
Private Bag 1421, East London, 5200, South Africa
+27 43 7085407

ssemugab@wsu.ac.za

Ruth de Villiers
School of Computing
University of South Africa
P O Box 392, Unisa, 0003, South Africa
+27 12 429 6559

dvillmr@unisa.ac.za

ABSTRACT

Usability evaluation of e-learning applications is a maturing area, which addresses interfaces, usability and interaction from human-computer interaction (HCI) and pedagogy and learning from education. The selection of usability evaluation methods (UEMs) to determine usability problems is influenced by time, cost, efficiency, effectiveness, and ease of application. Heuristic evaluation (HE) involves evaluation by experts with expertise in the domain area and/or HCI. This comparative evaluation study investigates the extent to which HE identifies usability problems in a web-based learning application and compares the results with those of survey evaluations among end-users (learners). Severity rating was conducted on a consolidated set of usability problems and further comparison of findings was done on the major and minor problems. The results of HE correspond closely with those of the survey. However the four expert evaluators identified more problems than the 61 learners and identified 91% of the learners' problems, when major problems only were considered. HE by a competent and balanced set of experts showed itself to be an appropriate, efficient and effective UEM for e-learning applications.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentations**]: User Interfaces – *evaluation/methodology;*

K.3.1 [**Computers and Education**]: Computer Use in Education – *computer-assisted learning.*

General Terms

Design, Human Factors, Measurement

Keywords

E-learning, evaluation criteria, heuristic evaluation, human-computer interaction, questionnaire survey, usability evaluation, usability evaluation methods, web-based learning

1. INTRODUCTION

Usability is a key issue in human-computer interaction (HCI) since it is the aspect that commonly refers to quality of the user interface [30]. The International Standards Organisation (ISO)

defines usability as [16]: *The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context.* Usability evaluation is concerned with gathering information about the usability or potential usability of a system, in order to assess it or to improve its interface by identifying problems and suggesting improvements [35]. Various usability evaluation methods (UEMs) exist, e.g. analytical, expert heuristic evaluation, survey, observational, and experimental methods [6, 13, 35]. Studies have been done to compare UEMs, but the area is still relatively incomplete as a research topic and as an applied body of knowledge [14]. To evaluate the usability of a system and to determine usability problems, it is important to select an appropriate UEM/s [10, 37], taking cognisance of efficiency, time, cost-effectiveness, ease of application, and expertise of evaluators [13, 30]. The present study focuses on usability evaluation of e-learning applications. Usability of e-learning poses its own requirements, hence its usability evaluation is different from that of general task-oriented systems and requires different criteria. It is a maturing area, in which research has been conducted by, among others, [1], [7], [8], [15], [25], [36], [40] and [42]. Evaluation of e-learning should address aspects of pedagogy and learning from educational domains as well as HCI factors such as the effectiveness of interfaces and the quality of usability and interaction.

The purpose of this study is primarily to compare the results of evaluation by two different means and to a less extent to evaluate the target system. In our comparative evaluation research, an appropriate framework of evaluation criteria is proposed and used as the basis for heuristic evaluation by experts and survey evaluation among end users (learners) on the same e-learning application at a tertiary institution. Section 2 briefly outlines heuristic evaluation. Section 3 sets out the research design and methodology of this study. The components of the research design include identification of the target application and evaluation criteria, conducting of the evaluations, and comparison and discussion of findings. These are discussed in depth in Sections 4 to 8, while Section 9 closes the paper.

2. HEURISTIC EVALUATION AND ITS APPLICATION IN E-LEARNING

Heuristic evaluation (HE) is a usability inspection technique originated by Nielsen [26, 27], in which a small set of expert evaluators, guided by a set of usability principles known as heuristics, determine whether a system conforms to these and

identify specific usability problems in the system. It is the most widely used UEM for computer system interfaces. It is described as fast, inexpensive, and easy to perform, and can result in major improvements to user interfaces [4, 5, 19, 24]. HE is often done during development, but can be very effective when used on real, operational systems [26, 31] as in this study, so as to compare the results of evaluation by two different methods on exactly the same application.

Over the past decade there has been widespread use of web-based e-learning applications for distance and classroom learning [41], yet little has been done to critically examine their usability [43]. Usability features are frequently not considered in development, sometimes because instructors and courseware developers are not trained to do so or because they lack the technological skills [41]. However, a main reason for the oversight is that usability evaluation can be difficult, time consuming and expensive [20]. If, in addition to its attributes of low cost and relative simplicity, HE is shown to be effective, efficient, and sufficient to identify the problems that impede learners, it would qualify as an ideal method for evaluating web-based e-learning. This study investigates the extent to which HE identifies usability problems in a web-based e-learning application.

3. RESEARCH DESIGN AND METHODOLOGY

3.1 Research Question

With the issues raised in Sections 1 and 2 as a background, the following research question emerges:

*To what extent do the findings of heuristic evaluation by experts of a web-based e-learning application correspond with the findings of survey evaluation by learners (*in the context of a tertiary institution*)?*

3.2 Design and Methodology

The research design involves a form of case study, namely a comparative evaluation study on a web-based learning (WBL) environment built for a specific course in a particular institution. We investigate and compare the findings of two evaluations by two different UEMs. A case study is defined as an investigation to answer specific research questions, requiring a range of different evidence from the case setting, which is abstracted and collated to get answers to the research question [11]. Submethods such as interviews and observation can be used to collect multiple evidences, i.e. triangulation [11]. In this study more than one method was applied in each form of evaluation of the target system. The main components of the research design were:

1. Identification of the target application.

2. Generation of a set of criteria/heuristics suitable for usability evaluation of e-learning applications, in particular, WBL environments.

3. End-user evaluations by means of criterion-based learner surveys on the target application using questionnaires and a focus group interview.

4. An heuristic evaluation with a competent and complementary set of experts, using the synthesised set of criteria as heuristics, followed by severity rating on the consolidated set of problems.

5. Data analysis to answer the research question, investigating in particular how the usability problems identified by heuristic evaluation compare with those identified by end users in the learner-surveys.

4. THE TARGET APPLICATION

The evaluations were conducted on *Info3Net*, the course website for Information Systems 3 at Walter Sisulu University (WSU) situated in East London in the Eastern Cape. Information Systems (IS) 3 covers *Advanced Databases* in the first semester and *Project Management* in the second. The *Info3Net* WBL site was selected as the target application to be evaluated, due to its convenience, accessibility and ease of control. It had been designed by the first author (lecturer and subject co-ordinator of IS 3) to address the database part of the course, and used for the first time in 2005.

WBL environments can be set up to supplement face-to-face teaching and learning, or as a stand-alone tool for instruction [17]. In this case, the former approach, known as blended learning, was used. About 80 students in the final year of diplomas in Information Technology or Financial Information Systems used *Info3Net* to supplement the traditional contact sessions. They all had a sound technological background and were competent in computing.

5. THE USABILITY EVALUATION CRITERIA AND GENERAL BACKGROUND

5.1 Evaluation Criteria/Heuristics

As noted in Section 1, evaluation criteria/heuristics (terms used interchangeably) for e-learning should address interfaces, usability and interaction from HCI, as well as pedagogy and learning from education. A study was undertaken by the authors [38] to establish an appropriate set of 20 criteria within three categories for evaluating WBL applications. This multi-faceted framework (see Table 1) supports Hosie's and Schibeci's concept [15] of 'context-bound'/'context-related' evaluation. Nielsen's [27] heuristics form the basis of Category 1, extended to customise them for educational purposes by using Squires' and Preece's [36] 'learning with software' heuristics. Other sources used in generating the criteria were, [1], [18], [19], [23], [34], [35] and [39]. Each criterion has a list of associated sub-criteria or guidelines to help evaluators assess *Info3Net.* These are also shown in Table 1, but they can be customised to other contexts. The expert evaluations described in Sections 7 excludes five criteria (Criteria 12, 15, 16, 17 & 19) regarding personal learning experiences, which are specifically for learners/students (terms used interchangeably) but not for expert evaluators, so that 15 of the 20 criteria were common to both studies.

Table 1. Evaluation criteria for web-based learning - a framework

Category 1: General interface usability criteria (based on Nielsen's heuristics, modified for e-learning context)	
1	**Visibility of system status** • The website keeps the user informed about what is going on through constructive, appropriate and timely feedback. • The system responds to user-initiated actions. There are no surprise actions by the site or tedious data entry sequences.
2	**Match between the system and the real world i.e. match between designer model and user model** • Language usage such as terms, phrases, symbols, and concepts, is similar to that of users in their day-to-day environment. • Metaphor usage corresponds to real-world objects/concepts, e.g. understandable and meaningful symbolic representations are used to ensure that the symbols, icons and names used are intuitive within the context of the performed task. • Information is arranged in a natural and logical order.
3	**Learner control and freedom** • Users control the system. • Users can exit the system at any time even when they have made mistakes. • There are facilities for Undo and Redo.
4	**Consistency and adherence to standards** • The same concepts, words, symbols, situations, or actions refer to the same thing. • Common platform standards are followed.
5	**Error prevention, in particular, prevention of peripheral usability-related errors** [36] • The system is designed such that the users cannot easily make serious errors. • When a user makes an error, the application gives an appropriate error message.
6	**Recognition rather than recall** • Objects to be manipulated, options for selection, and actions to be taken are visible. • The user does not need to recall information from one part of a dialogue to another. • Instructions on how to use the system are visible or easily retrievable whenever appropriate. • Displays are simple and multiple page displays are minimised.
7	**Flexibility and efficiency of use** • The site caters for different levels of users, from novice to experts. • Shortcuts or accelerators, unseen by novice users, are provided to speed up interaction and task completion by frequent users. • The system is flexible to enable users to adjust settings to suit themselves, i.e. to customise the system.
8	**Aesthetics and minimalism in design** • Site dialogues do not contain irrelevant or rarely needed information, which could distract users as they perform tasks.
9	**Recognition, diagnosis, and recovery from errors** • Error messages are expressed in plain language. • Error messages define problems precisely and give quick, simple, constructive, specific instructions for recovery. • If a typed command results in an error, users need not retype the entire command, but repair only the faulty part.
10	**Help and documentation** • The site has a help facility and other documentation to support users' needs. • Information in these facilities is easy to search, task-focused, and lists concrete steps to accomplish a task.
Category 2: Website-specific criteria for educational websites	
11	**Simplicity of site navigation, organisation and structure** • The site has a simple navigational structure. • Users should know where they are and have options of where to go next, e.g. via a site map or breadcrumbs. • The navigational options are limited so as not to overwhelm the user. • Related information is placed together. • Information is organised hierarchically, moving from the general to the specific. • Common browser standards are followed. • Each page has all the required navigation buttons or hyperlinks (links), such as previous (back) next and home.
12	**Relevance of site content to the learner and the learning process** • Content is engaging, relevant, appropriate and clear to learners using the WBL site.

	• The material has no biases such as racial and gender biases, which may be deemed offensive. • It is clear which materials are copyrighted and which are not. • The authors of the content are of reputable authority.
	Category 3: Learner-centred instructional design, grounded in learning theory, aiming for effective learning
13	**Clarity of goals, objectives and outcomes** • There are clear goals, objectives and outcomes for learning encounters. • The reason for inclusion of each page or document on the site is clear.
14	**Effectiveness of collaborative learning (where such is available)** • Facilities and activities are available that encourage learner-learner and learner-teacher interactions. • Facilities for both asynchronous and synchronous communication, such as e-mail, discussion forums and chat rooms.
15	**Level of learner control** • Apart from controlling the interactions with the site, learners have some freedom to direct their learning, either individually or through collaborative experiences, and to have a sense of ownership of their learning. • Learners are given some control of the content they learn, how it is learned, and the sequence of units. • Individual learners can customise the site to suit their personal learning strategies. • Educators can customise learning artefacts to the individual learner, for example, tests and performance evaluations can be customised to the learner's ability. • Where appropriate, learners take the initiative regarding the methods, time, place, content, and sequence of learning
16	**Support for personally significant approaches to learning** • There are multiple representations and varying views of learning artefacts and tasks. • The site supports different strategies for learning and indicates clearly which styles it supports. • The site is used in combination with other mediums of instruction to support learning. • Metacognition (the ability of a learner to plan, monitor and evaluate his/her own cognitive skills) is encouraged. • Learning activities are scaffolded by learner support and by optional additional information.
17	**Cognitive error recognition, diagnosis and recovery** • Cognitive conflict, bridging and problem-based learning strategies are used in the recognition-diagnosis-recovery cycle. • Learners have access to a rich and complex environment where they can explore different solutions to problems. • Learners are permitted to learn by their mistakes and are provided with help to recover from cognitive errors. • Learners are given opportunities to develop personal problem-solving strategies.
18	**Feedback, guidance and assessment** • Apart from the system's interface-feedback by the system, considered under Criterion 1, learners give and receive prompt and frequent feedback about their activities and the knowledge being constructed. • Learners are guided as they perform tasks. • Quantitative feedback, e.g grading of learners' activities, is given, so that learners are aware of their level of performance.
19	**Context meaningful to domain and learner** • Knowledge is presented within a meaningful and authentic context that supports effective learning. • Authentic, contextualised tasks are undertaken rather than abstract instruction. • The application enables context- and content-dependent knowledge construction. • Learning occurs in a context of use so that knowledge and skills learned will be transferable to similar contexts. • The representations are understandable and meaningful, ensuring that symbols, icons and names used are intuitive within the context of the learning task.
20	**Learner motivation, creativity and active learning** • The site has content and interactive features that attract, motivate and retain learners, and that promote creativity on the part of learners, e.g. the online activities are situated in real-world practice, and interest and engage the learners. • To promote active learning and critical thinking, tasks require learners to compare, analyse and classify information, and to make deductions.

5.2 Ethical Aspects and Social Responsibility

Evaluations should be performed in a professional and ethical manner. Participants' rights must be protected [33]. End user or expert evaluators should sign a consent form and receive information documents, explaining the purpose, and assuring anonymity and that findings will be used for stated purposes only [33, 35]. These evaluations were conducted in an ethical and socially responsible manner. Participation was voluntary and participants could withdraw at any time. Students received no reward for participating. Expert evaluators were not paid for their professional expertise, but received a token gift in gratitude. Furthermore, where questionnaires are used, they

should be pilot tested to avoid any potential misunderstanding. This was done in the study, as described in Section 6.

6. SURVEY EVALUATIONS AMONG END-USERS (LEARNERS)

Query techniques, such as questionnaires and interviews aim to identify usability problems by asking users directly [9]. The questionnaire was designed using Gillham's principles [12]. The first section elicited basic demographic information and details of the respondents' experience. The main section was based on the 20 criteria of Section 5.1. For each criterion, straightforward, focused, single-issue statements, based on the sub-criteria, were generated to expand the meaning. Students rated *Info3Net* using the criteria, but a main aim was for them to use each criterion to name problems they had experienced in that regard, and write them in the space below. See an extract in Table 2 relating to Criterion 1 of Category 1. The last section of the questionnaire required an overall rating for *Info3Net* in relation to its usability. Questionnaires should be reviewed and tested with a small sample, according to [35], [29] and [12], who recommend a pilot study to avoid potential mis-understanding and cases where respondents do understand instructions/questions, but lack the background to respond. Our questionnaire was pilot tested by five students. Feedback from this study and observation by a researcher solved several problems that emerged.

Eighty registered students had used *Info3Net* during the semester, 61 of whom took part in the questionnaire survey, which was conducted on a single day in the usual laboratory setting, but in two separate groups. Sixty four problems were identified. To eliminate duplicates, a researcher meticulously considered them and combined those that were closely related. During consolidation, some were rephrased to correspond with terminology used by expert evaluators (7.4). For further data, the questionnaire was followed by a focus group interview with eight students, as advocated by [35]. This clarified and elaborated problems identified by the questionnaire. More emerged, resulting in a final list of 55 problems. Table 3 lists the problems that emerged under Criterion 3 of Category 1. The number of problems identified under each criterion is shown in Table 6 in Section 8.1, where the results are discussed and compared with those of the heuristic evaluation by experts.

7. HEURISTIC EVALUATION BY EXPERTS

Heuristic evaluation (HE) is an inspection technique whereby a set of experts evaluate whether a user interface conforms to defined usability principles, termed *heuristics* [9, 26, 33]. It is seldom possible for a single evaluator to identify all the usability problems in a system. Nielsen [27] recommends that a set of

Table 2. Extract from the Criterion 1 section of the questionnaire

1. Visibility of system status

1.1 The system keeps me informed through feedback about what is going on.

Strongly agree	Agree	Maybe	Disagree	Strongly disagree

1.2 I understand what the feedback means.

Strongly agree	Agree	Maybe	Disagree	Strongly disagree

1.3 I get feedback within a reasonable time.

Strongly agree	Agree	Maybe	Disagree	Strongly disagree

1.4 For every action I make, I can see or hear the results of that action.

Strongly agree	Agree	Maybe	Disagree	Strongly disagree

1.5 The system does not react in a manner that surprises me and it does not do anything unexpected.

Strongly agree	Agree	Maybe	Disagree	Strongly disagree

1.6 Write down any particular problem(s) you found in *Info3Net* in relation to this section.
_____

(*Actual open-ended response area was much larger – reduced for this paper.*)

Table 3. Problems identified by students with regard to Criterion 3

3. Learner control and freedom
There are no facilities for Undo and Redo.
The system is slow to respond.
There is no system exit button.
When doing a quiz/test, if the test is submitted before the time expires, one should have a chance to change answers within the time limit.
It is not easy to print site content.

three to five evaluators be used to identify 65-75% of the usability problems. Our approach is based on Nielsen's, using our custom-designed heuristics for web-based learning. The HE was supplemented by a severity rating process done by the experts to rank the final integrated set of problems ranked according to level of seriousness.

7.1 Identifying and Defining the Heuristics

The heuristics used were 15 of the 20 used in the end-user survey – see Section 5.1 and the results of the two evaluations are compared with regard to those 15. Each heuristic was divided into sub-criteria, as shown in Table 1.

7.2 Selection of Evaluators

Factors involved in selecting and inviting a balanced set of experts, are the number to use and their respective back-grounds. Nielsen's [27] cost-benefit analysis demonstrated optimal value with three or four evaluators. Despite this, the debate continues. [3] used eleven experts to assess usability of a university Web portal. Law and Hvannberg [22] reject the 'magic five assumption' and in the context of usability testing, used eleven participants to define 80% of the detectable usability problems. However, [19] argue that two to three evaluators who are experts both in the domain area and in HCI, so-called 'double experts', will point out the same number of usability problems as three to five 'single experts'. Furthermore, in a heuristic evaluation of an educational multimedia application by Albion [1], four evaluators were selected, with expertise in user interface design, instructional/educational design, and teaching. The same approach was followed in this study. Four expert evaluators with expertise in user interface design, instructional/educational design, and teaching, were invited and agreed to participate. Two are lecturers in the subject-matter domain and are also HCI experts, familiar with HE; these two can be classified as 'double experts'. Table 4 shows the profiles of the four, all of whom are academics who lecture, do postgraduate supervision, and conduct research. They come from three different tertiary institutions in different cities.

7.3 Briefing the Evaluators

Evaluators were briefed in advance about the HE process for the study, the domain of use of the target system, and the task scenarios to work through as advocated by [23] and [27]. In addition to the consent form (see 5.2) and a request to familiarise themselves with the heuristics before doing the actual evaluation, each evaluator was given a set of documents regarding the:

- *Phases of the HE:* Overall guidelines on the entire process up to severity rating; descriptions of the stages, the documents to be read, and the estimated time required to perform the evaluation.

- *System and user profile*: A general background to *Info3Net*, its learners, and how they used the application.

- *Procedure*: The procedure to be followed in conducting the HE. It included log-in details, details of the scenarios to perform, instructions on how to do the actual evaluation and how to compile the report of usability problems.

7.4 Actual Evaluation

As advised by Nielsen, each of the four expert evaluators conducted his/her evaluation independently. They performed the same tasks as those in the end-user survey evaluations. Apart from issues arising out of these tasks, experts were asked to assess general features of the application, just as the learners did. They used the 15 selected criteria (see 5.1) along with their sub-criteria. They were given a three-week window during which to find a two-hour slot to evaluate *Info3Net* on the Internet. Due to a high-pressure academic period, only one of the four did so within the specified period, and spent two hours on it. In discussions with the others, it was decided that the researcher would be available as an objective observer-facilitator during the evaluation process, *not to intervene*, but to expedite the process by answering queries and serving as a scribe. This enables the evaluator to concentrate on pointing out usability problems, while the researcher clarifies issues that may arise [23, 27]. With this support, the other three experts performed their evaluations in the next ten days, taking about an hour to hour-and-a-quarter to do so. A list of 77 unique problems was initially identified by the set of experts.

Those that were closely related were consolidated and combined to eliminate duplicates, as described in Section 6 regarding learners' problems. The final list of problems from the expert evaluation comprised 58 problems. The number of problems identified by each expert varied from 9 to 31 with an average of 20.5. There was a tendency for different evaluators to concentrate on different heuristics, highlighting their complementary strengths. Table 6 in 8.1 shows the number of problems for each criterion, comparing the results of the experts and the end-user. Once all evaluations were complete, severity rating was conducted on the consolidated set of usability problems resulting from end-user surveys and the expert evaluations.

Table 4. Profiles of the expert evaluators

Evaluators	Evaluator 1	Evaluator 2	Evaluator 3	Evaluator 4
Highest qualifications	Med; PhD	Mcom (IS)	Med; MSc(IS)	Med; DSc(Ed)
Professional role	Prof. in Dept. of Education.	Senior lecturer in IS Dept. Course manager: IS Hons.	Prof. in IS Dept; Course manager: IS masters.	Senior lecturer in Dept. of IT.
Duties/courses taught (relevant to this study)	ICT, Computer Literacy & Psychology of Education	HCI at honours level & Database Design at 3rd year level (same level as that of *Info3Net*).	HCI at Honours level and Database Design at 3rd year level.	Research Skills for BTech and IT Internet; programming at 2nd year level

7.5 Severity Rating

Severity rating, i.e assigning relative severities to individual problems, can be performed to determine each problem's level of seriousness. It is usually estimated on a 3- or 5-point Likert scale. The experts can either do ratings during their HEs or later, when all the problems have been aggregated. The latter approach is advantageous [1, 23, 24] and is used in the present study. Table 5 shows the 5-point rating scale [32] used to assess the problems and assign severities. The final option indicates a non-problem. The scale is similar to that used by Nielson [27] with an additional rating item, 'Medium' between the 'Major' and 'Minor' problem ratings.

The problems identified by the four evaluators (58 problems) were merged with those pinpointed by the learners (55 problems) to make a single list of all the identified problems. This final consolidated set comprised 75 problems, categorised according to the criterion they violated. This set was sent to the expert evaluators for severity rating. To provide more information, each problem was accompanied by a weighting indicating how many experts and how many students had identified it. The problems for each criterion were grouped in descending order, according to the number of experts that had identified it. All four expert evaluators e-mailed back the completed form within three days of receiving it.

8. COMPARISON OF THE EVALUATIONS AND DISCUSSION OF FINDINGS

8.1 Overall Comparison of Results

Table 6 shows the number of problems identified by experts and students for each criterion, as well as the number of problems common to both evaluations. In total, 75 problems were identified. Of these, 58 were identified by experts and 55 by learners, as previously mentioned, and 38 by both groups. This means that the experts identified 77% of all the problems and the learners 73%. Although the number of learners, namely 61, who participated in the evaluation, was significantly more than the number of experts, namely 4, the former large group identified fewer problems than the latter very small group. This is in line with Nielsen's claim that end users are not good at identifying usability problems, since they have a 'poor conceptual understanding of usability' [27] and possibly also of learning principles. However, since the experts identified approximately three-quarters of the problems (77%), the present findings support the statement by [28] that experts are not perfect in identifying problems.

Table 6 also shows the interesting result that more than half, 51% (38 out of 75), of all problems were identified by both groups. Furthermore, the students identified 38 of 58 (66%) of the problems identified by experts. Similarly, the experts identified 38 of 55 (69%) of the problems identified by learners. These results demonstrate a notable correspondence between the results of the two evaluations.

Table 7 shows that close correspondence also occurred in the number of problems found per category per group, although the percentages of common problems in the three categories (General, Web, Educational): 53%, 20% and 63% respectively, show that certain different problems were noted by the two groups. There were ten, only one, and five criteria in the three categories respectively, possibly contributing to the low level of common findings in the Web category. However, the differences also emphasise the synergistic value of combining different UEMs.

This study is based on the premise in Section 2, that HE is inexpensive, easy, and fast to perform when compared to other evaluation methods. Issues of cost and researcher-time were not formally investigated in this study. However, the researchers spent considerably less time and effort on the heuristic evaluation than on the survey with respect to preparation of materials, conducting the evaluations, and analysing the data. For example, only four sets of evaluation results were processed in the HE, versus 61 sets in the survey yet, due to their experience and expertise, the four experts identified 77% of the set of combined problems and 69% of the problems identified by the group of learners.

8.2 Severity of Problems

Section 7.4 explains that each of the four expert evaluators independently rated the severity of the problems in the combined list of 75 problems, as advocated by leading researchers [1, 23, 24, **27**]. The rating scale of 1 to 5 represents cosmetic, minor, medium, major, and catastrophic problems, respectively. The mean of the average ratings is 3.0 with a standard deviation of 0.8, indicating that most problems were considered to be medium problems (Table 5), to which users can adapt quickly and still use *Info3Net* satisfactorily. The mean scores for Evaluators 1 to 4 were 3.1, 3.2, 2.9 and 2.8 respectively, each of which is close to the overall mean of 3.0. A close look at the source data shows that most problems were allocated very similar scores by all the expert evaluators. Moreover, the experts were not biased in assigning ratings, since each assigned high ratings to certain problems they had not recognised themselves. In so doing, they acknowledged and

Table 5. Five-point rating scale for severity of usability problems [32]

Cosmetic problem: will not affect the use of the system. Fix it if possible.	1
Minor problem: users can easily work around the problem. Fixing it should be given a low priority.	2
Medium problem: users are likely to encounter this problem but will quickly adapt. Fixing it should be a medium priority.	3
Major problem: users will find this problem difficult but may work around it. Fixing it should have a high priority.	4
Catastrophic problem: users will be unable to do their work because of this problem. Fixing it is mandatory.	5
Not Applicable: I don't consider this to be a problem.	N

Table 6. Number of problems identified by students and by experts

Category	Criterion	Number of problems identified			
		Experts	Students	Common	All
Category 1: General interface design (49 problems)	1	4	2	1	5
	2	6	3	3	6
	3	5	5	4	6
	4	3	5	3	5
	5	2	3	2	3
	6	5	2	2	5
	7	6	6	5	7
	8	3	4	3	4
	9	1	3	1	3
	10	3	4	2	5
Category 2: Educational website-specific aspects (10 problems)	11	7	5	2	10
Category 3: Learner-centred instructional design (16 problems)	13	4	3	3	4
	14	1	2	1	2
	18	6	6	4	8
	20	2	2	2	2
Total (all problems)		**58**	**55**	**38**	**75**
% of all 75 problems		77%	73%	51%	100%
% of other group's problems		69%	66%		

Table 7. Problems identified, according to category

Category	Experts		Students		Common		All groups	
	No of prob's	% of prob's in cat	No of prob's	% of prob's in cat	No of prob's	% of prob's in cat	No of prob's in cat	% of total prob's
General	38	78%	37	76%	26	53%	49	65%
Web	7	70%	5	50%	2	20%	10	13%
Educational	13	81%	13	81%	10	63%	16	22%
All problems identified by group	58	100%	55	100%	38	100%	75	100%

validated the problems identified by their peers with different specialities. This ratifies the value of complementary strengths and indicates the benefit of severity rating after evaluations, using a consolidated set of problems generated by a balanced set of experts as well as those pinpointed by end users.

8.3 Major and Minor Problems

As stated, the purpose of severity rating is to determine the seriousness of usability problems and categorise them accordingly. Problems can be major or minor. Data was extracted from the results of the experts' ratings in such a way that those problems with an average of 4 to 5 (major to catastrophic) are viewed as *major* problems and those with an average of 1 to 2 (cosmetic to minor) as *minor* problems. After this categorisation of problems, further comparison of the findings of experts and students was done. Table 8 shows the number of problems in the major and minor categories and their corresponding identification percentages. Twelve problems were

thus found to be major and fourteen are minor. Of the final 75 problems, 49 problems with average ratings between 2 and 5 are excluded from these sets, i.e. they are neither major nor minor. Table 8 shows that of the twelve major problems, eleven were identified by the experts, eleven by the learners, and ten problems by both groups. This corresponds to 92%, 92% and 83% as shown in Table 8. Secondly, since ten problems were common, it means that of the problems in the major category, ten of the eleven identified by learners, i.e. 91%, were also identified by the experts, compared to 69% (38 of 55) of the learners' problems that were identified when all the problems were considered (see Table 6 and discussion). This indicates that both sets of evaluators, experts and learners, identified a high percentage, 92%, of major problems, and that there was consensus in this respect. In fact, out of the twelve problems in the major category, only one problem was identified by experts but not by learners, and one by learners but not by experts.

When considering minor problems, no such close correspondence is evident. Table 8 shows that of the fourteen minor problems, the experts identified eight (64%) of them, learners five (43%) and only one (7%) was common to both groups. This data shows that both sets of evaluators identified a low percentage of minor problems and that there was little agreement, only 7%, on this. These results support similar studies [21, 27], which show that major problems are easier to find than minor ones. To summarise, in this study both sets of evaluators identified 92% of the major problems, compared to 64% by experts and 43% by learners of the minor problems. [27] found that, although major usability problems are easier to identify, when all problems are considered, the number of minor problems will exceed the number of major ones. This occurrence is supported in the present study.

9. CONCLUSION AND FUTURE RESEARCH

The research reported in this paper can be described as 'meta-evaluation' in that it is focused not primarily on the target system, but rather on investigating two UEMs, in particular, heuristic evaluation, studying the extent to which HE by a small set of appropriate experts, can identify usability problems in a web-based learning application. This was done by comparing the HE results with those of an end-user survey among the learner population, using evaluation criteria custom-generated for e-learning applications.

Although the learners had been using *Info3Net* for some time, they identified a slightly lower proportion of the total problems, 73%, than the 77% of the expert evaluators, who were having their first exposure. This is in line with Nielsen's [27] claim (see 8.1) that end users are not as good as experts in identifying usability problems. When major problems only were considered, the proportion of problems identified rose to the same value for both groups, namely 92%. The increase in percentage shows that it is easier, for both learners and experts, to identify major problems. A further notable finding is that the four experts were able to identify 69% of the problems identified by the learners, while learners identified 66% of the experts' problems. The percentage of learner-problems also identified by experts rose to 91% when major problems only were considered.

We now return to the research question posed in Section 3.1, namely:

To what extent do the findings of heuristic evaluation by experts of a web-based e-learning application correspond with the findings of survey evaluation by learners?

The conclusion from these findings of a comparative evaluation study, is that the results of heuristic evaluation by experts correspond closely with those of survey evaluation among end-users (learners). In fact, the HE results are better than the survey results (see Table 6). They were produced by only four experts compared to 61 learners, and the experts were experiencing their first exposure to *Info3Net*, whereas the learners had used it for a semester. Yet in total the experts identified 58 problems compared to the learners' 55, and the percentage of all problems identified by HE was 77%, compared to 73% by the learner survey. The experts identified 69% of the learner-identified problems and 91% of the learner-identified problems when major problems only are considered.

The findings of this study indicate that heuristic evaluation, if conducted by a competent and complementary group of experts, is an appropriate, efficient, and highly effective usability evaluation method in the context of e-learning, as well as relatively easy to conduct and inexpensive. The researchers recommend that HE should, ideally, be supplemented with methods where users themselves identify usability or learning problems. This is in line with proposals that reliable evaluation can be achieved by systematically combining inspection with user-based methods [**2**, 7]. In cases where only one approach has to be selected, the findings of this research can be used to propose heuristic evaluation as the optimal method.

A valuable secondary benefit of the study is that the problems identified in *Info3Net* can be addressed in future upgrades. Another major contribution of this research is the generation of the framework of evaluation criteria, which is transferable to

Table 8. Problems identified by severity rating as major and minor problems

Level of severity	Experts	Students	Common	All groups
Major problems	11	11	10	12
% of major problems identified	92%	92%	83%	100%
Minor problems	9	6	1	14
% of minor problems identified	64%	43%	7%	100%

other e-learning contexts, where the sub-criteria can be customised to the particular environment or system. Recommendations for further research include the following:

- Re-designing *Info3Net* in an action research approach, to solve some of the problems identified. The application could be re-evaluated to determine the impact of the changes.

- Applying the evaluation criteria generated in this study to evaluate other web-based e-learning applications.

- Adapting the framework of criteria to customise them for other contexts or other forms of e-learning.

- Determining which criteria are suitable for application by educators only and which by learners only.

10. ACKNOWLEDGEMENT

11. References

[1] Albion, P.R. 1999. *Heuristic Evaluation of Multimedia: From Theory to Practice.* http://www.usq.edu.au/users/albion/papers/ascilite99.html.

[2] Ardito, C., Costabile, M.F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T. & Rossano, V. 2006. An Approach to Usability Evaluation of E-Learning Applications. *Universal Access in the Information Society*, 4(3): 270-283.

[3] Avouris, N., Tselios, N., Fida, C. & Papachristos, E. 2003. Website Evaluation: A usability-Based Perspective. In: Y. Manolopoulos, S. Evripidou & A.C. Kakas. (Eds), *Lecture Notes in Computer Science. Vol 2563: Advances in Informatics, 8th Panhellenic Conference on Informatics*: 217-232. Berlin: Springer-Verlag.

[4] Belkhiter, N., Boulet, M., Baffoun, S. & Dupuis, C. 2003. Usability Inspection of the ECONOF System's User Interface Visualization Component. In: C. Ghaoui. (Ed.), *Usability Evaluation of Online Learning Programs*. Hershey, P.A.: Information Science Publishing.

[5] Blandford, A., Keith, S., Connell, I. & Edwards, H. 2004. Analytical Usability Evaluation for Digital Libraries: A Case Study. In: *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries 2004*: 27-36. Tucson: ACM Press.

[6] Brinck & Wood. 2002. *Usability for the Web: Designing Web Sites that Work*. San Francisco: Morgan Kaufmann Publishers, (2002).

[7] Costabile, M.F, De Marsico, M., Lanzilotti, R., Plantamura, V.L. & Roselli, T. 2005. On the Usability Evaluation of E-Learning Applications. In: *Proceedings of the 38th Hawaii International Conference on System Science*: 1-10. Washington: IEEE Computer Society.

[8] De Villiers, R. 2006. Multi-Method Evaluations: Case Studies of an Interactive Tutorial and Practice System. In: *Proceedings of InSITE 2006 Conference*. Manchester, United Kingdom.

[9] Dix, A.J., Finlay, J.E., Abowd, G.D. & Beale, R. 2004. *Human-Computer Interaction*. 3rd Ed. Harlow Assex: Pearson Education Limited.

[10] Fitzpatrick, R. 1999. *Strategies for Evaluation of Software Usability*. http://www.comp.dit.ie/rfitzpatrick/papers/chi99%20strategies.pdf.

[11] Gillham, B. 2000. *Case Study Research Methods*. London: Bill Gillham.

[12] Gillham, B. 2000. *Developing a Questionnaire*. London: Bill Gillham.

[13] Gray, W.D. & Salzman, M.C. 1998. Repairing Damaged Merchandise: A Rejoinder. *Human-Computer Interaction* 13 (3): 325-335.

[14] Hartson, H.R., Andre, T.S. & Williges, R.C. 2003. Criteria for Evaluating Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 15(1): 145-181.

[15] Hosie, P. & Schibeci, R. 2001. Evaluating Courseware: A need for more context bound evaluations? *Australian Educational Computing*, 16(2):18-26.

[16] ISO. 1998. *ISO-9241: Guidance on Usability Standards*. http://www.iso.ch/iso/en/CatalogueListPage.CatalogueList?ICS1=13&ICS2=180.

[17] Jackson,R.H. 2004. *Weblearning Resources*. hhttp://www.knowledgeability.biz/weblearning/default.htm.

[18] Jones, A., Scanlon, E., Tosunoglu, C., Morris, E., Ross, S., Butcher, P. & Greenberg, J. 1999. Contexts for Evaluating Educational Software. *Interacting with Computers*, 11(5): 499-516.

[19] Karoulis, A. & Pombortsis, A. 2003. Heuristic Evaluation of Web-Based ODL Programs. In: C. Ghaoui. (Ed.), *Usability Evaluation of Online Learning Programs*. Hershey, P.A.: Information Science Publishing.

[20] Kjeldskov, J., Skov, M.B. & Stage, J. 2004. Instant Data Analysis: Conducting Usability Evaluations in a Day. In: *Proceedings of the third Nordic Conference on Human-Computer Interaction*: 233-240. Tampere: ACM International Conference Proceeding Series.

[21] Law, L.C. & Hvannberg, E.T. 2002. Complementarity and Convergence of Heuristic Evaluation and Usability Test: A Case Study of Universal Brokerage Platform. *Proceedings of the Second Nordic Conference on Human-computer Interaction*: 71-80. Arhus: ACM International Conference Proceeding Series.

[22] Law, L.C. & Hvannberg, E.T. 2004. Analysis of combinatorical user effect in international usability tests. *Proceedings of* CHI2004: 9-16.

[23] Levi, M.D. & Conrad, F.G. 1996. A Heuristic Evaluation of a World Wide Web Prototype. *Interactions*, 3(4): 50-61.

[24] Lindgaard, G. 2004. *Are the Notions of Thoroughness, Efficiency, and Validity Valid in HCI Practice?*. http://www.carleton.ca/hotlab/hottopics/Articles/Gitte_usability.html.

[25] Masemola, S.S. & De Villiers, M.R. 2006. Towards a Framework for Usability Testing of Interactive E-Learning Applications in Cognitive Domains, Illustrated by a Case Study. In: J. Bishop & D. Kourie. *Service-Oriented Software and Systems. Proceeding of SAICSIT 2006*: 187-197. ACM International Conference Proceedings Series.

[26] Nielsen, J. 1992. Finding Usability Problems through Heuristic Evaluation. In: *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*: 373-380. Monterey: ACM Press.

[27] Nielsen, J. 1994. Heuristic Evaluations. In: J. Nielsen & R.L. Mack. (Eds), *Usability Inspection Methods*. New York: John Wiley & Sons.

[28] Nielsen, J. & Molich, R. 1990. Heuristic Evaluation of User Interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People*: 249-256. Seattle: ACM Press.

[29] Olivier, M.S. 2003. *Information Technology Research*. Johannesburg: Van Schaik.

[30] Parlangeli, O., Marchingiani, E. & Bagnara, S. 1999. Multimedia in Distance Education: Effects of Usability on Learning. *Interacting with Computers*, 12(1): 37-49.

[31] Peng, L.K., Ramaiach, C.K, & Foo, S. 2004. Heuristic-Based User Interface Evaluation at Nanyang Technological University in Singapore. *Program: Electronic Library and Information Systems*, 38 (1): 42-59.

[32] Pierotti, D. 1996. *Usability Techniques: Heuristic Evaluation Activities*. http://www.stcsig.org/usability/topics/articles/he-activities.html.

[33] Preece, J., Rogers, Y. & Sharp, H. 2002. *Interaction Design: Beyond Human-Computer Interaction*. New York: John Wiley & Sons.

[34] Reeves, T.C. & Reeves, P.M. 1997. Effective Dimensions of Interactive Learning on the World Wide Web. In: B.H. Khan. (Ed.), *Web-Based Instruction*. New Jersey: Educational Technology Publications.

[35] Shneiderman, B. & Plaisant, C. 2005. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 4rd Ed. New York: Addison-Wesley.

[36] Squires, D. & Preece, J. 1999. Predicting Quality in Educational Software: Evaluating for Learning, Usability and the Synergy Between them. *Interacting with Computers,* 11(5): 467-483.

[37] Ssemugabi, S. 2006. *Usability Evaluation of a Web-Based E-Learning Application: A Study of Two Evaluation Methods*. MSc Dissertation, University of South Africa.

[38] Ssemugabi, S. & De Villiers, M.R. 2007. Usability and Learning: A Framework for Evaluation of Web-Based E-Learning Applications. In: *Proceedings of the ED-MEDIA 2007 - World Conference on Educational Multimedia, Hypermedia & Telecommunications*: 906-913. Vancouver, Canada.

[39] Storey, M.A, Phillips, B., Maczewski, M. & Wang, M. 2002. Evaluating the Usability of Web-Based Learning Tools. *Educational Technology & Society*, 5(3):91-100.

[40] Van Greunen, D. And Wesson, J L. 2002. Formal usability testing of interactive educational software: A case study. *World Computer Congress (WCC): Stream 9: Usability*. Montreal, Canada, August 2002.

[41] Vrasidas, C. 2004. Issues of Pedagogy and Design in E-learning System. In: *ACM Symposium on Online Learning*: 911-915. Nicosia: ACM Press.

[42] Wesson, J.L. & Cowley, N. L. 2003. The Challenge of Measuring E-Learning Quality: Some Ideas from HCI. *IFIP TC3/WG3.6 Working Conference on Quality Education @ a Distance*: 231-238. Geelong, Australia.

[43] Zaharias, P. 2006. A Usability Evaluation Method for E-Learning: Focus on Motivation to Learn. In: *CHI '06 Extended Abstracts on Human Factors in Computing Systems*: 1571-1576. Montreal: ACM Press.