

# Effectiveness of heuristic evaluation in usability evaluation of e-learning applications in higher education

Samuel Ssemugabi,<sup>1</sup> M R (Ruth) de Villiers<sup>2</sup>

School of Computing, University of South Africa, South Africa.

---

## ABSTRACT

The Internet, World Wide Web (WWW) and e-learning are contributing to new forms of teaching and learning. Such environments should be designed and evaluated in effective ways, considering both usability- and pedagogical issues. The selection of usability evaluation methods (UEMs) is influenced by the cost of a methods and its effectiveness in addressing users' issues. The issue of usability is vital in e-learning, where students cannot begin to learn unless they can first use the application. Heuristic evaluation (HE) remains the most widely-used usability evaluation method. This paper describes meta-evaluation research that investigated an HE of a web-based learning (WBL) application. The evaluations were based on a synthesised framework of criteria, related to usability and learning within WBL environments. HE was found to be effective in terms of the number and nature of problems identified in the target application by a complementary team of experienced experts. The findings correspond closely with those of a survey among learners.

## CATEGORIES AND SUBJECT DESCRIPTORS

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – *evaluation/methodology*;

K.3.1 [Computers and Education]: Computer Uses in Education – *computer-assisted instruction (CAI)*.

## KEYWORDS

Comparative evaluation, e-learning, evaluation criteria, heuristic evaluation, usability evaluation methods, user-based survey, web-based learning

---

## 1. INTRODUCTION

The advent of the Internet, the World Wide Web (WWW) technologies and e-learning in the last two decades has contributed to a new era of education [29, 54]. It is important that e-learning environments are designed and evaluated in an educationally effective manner by taking into account both usability and pedagogical issues [3, 6, 32, 47]. However, despite widespread use of e-learning, the critical examination of its usability is a newer field [56]. For example, several higher education institutions in South Africa have developed web-based learning (WBL) applications and tools without adequate consideration of usability [52]. Other studies [32, 57] show that although there are many reasons for high attrition from e-learning programs, such as irrelevant content and inappropriate use of technology, the major factor is poor usability of e-learning applications.

There are various usability evaluation methods (UEMs): analytical; inspection methods such as expert heuristic evaluation; surveys by questionnaires and interviews; observational; and experimental methods [10, 16, 40, 46]. Selection of an appropriate UEM requires consideration of its cost and effectiveness [3]. Heuristic evaluation (HE) is the most widely used UEM since it is inexpensive and easy to apply [5, 24]. However, according to Hollingsed and Novick,

overviewing 15 years of research and practice, 'The assessment of the effectiveness of heuristic evaluation continues as an active research thread' [17, p. 250]. In the case of e-learning, selection of a UEM is particularly important, because unless a system is easily usable, learning is obstructed and students spend more time learning how to use it than learning from it [3].

The research design of this paper involves a study based on using two UEMs in a complementary way, namely, HE supplemented with user surveys to evaluate a WBL application, taking both usability- and pedagogical aspects into account. The study aimed primarily to investigate the use of HE in the context of WBL. Section 2 is a literature survey that overviews the main themes of the research i.e. e-learning and usability evaluation. In Section 3, we describe the research design and methodology. This is followed by Section 4, which presents our framework of 20 custom-designed criteria, synthesized for usability evaluation of web-based educational environments. Thereafter, a description of the actual evaluation is provided in Section 5. In Section 6, we compare the results of the evaluations and discuss the findings. Finally Section 7 presents conclusions and notes the contributions of the study.

---

<sup>1</sup> Email: ssemus@unisa.ac.za,

<sup>2</sup> Dvillmr@unisa.ac.za

## 2. LITERATURE SURVEY

### 2.1 E-learning and WBL

E-learning is a form of teaching and learning that includes instruction delivered via a broad variety of electronic media and e-learning artifacts, including the Internet/intranets/extranets, web-based learning, satellite broadcasts, video/audio tape or DVD, multimedia CD-ROM, online instruction, and traditional computer-assisted learning (CAL) [7, 9, 13, 31, 55]. The teaching and learning may involve one or more learners at a distance, face-to-face or both [44]. WBL entails the use of learning materials that are delivered via a Web browser, including materials packaged on CD-ROM or any other storage media [25]. An important aspect of e-learning is the careful consideration of the underlying pedagogy and recognised learning theories, i.e. how learning takes place. E-learning can be effective only when there is alignment of the approaches to learning (learning theories) with the use of technology [13, 45]. Technology should be the medium and not the message [7].

### 2.2 Usability in the context of e-learning

There is a current focus on research on the intersection of human-computer interaction (HCI) and e-learning environments to determine how to engage learners and motivate them to interact with these systems [42]. Usability is a key issue in HCI, since it is the aspect that commonly refers to quality of the user interface [37]. The International Standards Organisation (ISO-9241) defines usability as [20]: *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 of e-learning applications significantly affects learning, since learner interactions with e-learning interfaces should result in true learning rather than in successful completion of tasks. As well as being a computing system, an e-learning product is also tutorial matter. The effectiveness of learning and users' satisfaction with a resource are therefore part of its usability. The interface should not be a barrier to learning or distract learners from achieving their learning goals. Ideally, the interface should be virtually invisible to the learner [3, 6, 53]. When systems are not easily usable, learners might spend excessive time trying to understand the system and how to use it, rather than engaging with the actual learning content [6]. In fact, there should be a synergy between the learning process and the interaction with the application [3, 47].

### 2.3 Usability evaluation of e-learning

Usability evaluation is concerned with gathering information about the usability or potential usability of a system, in order to assess it or improve its interface by identifying problems and suggesting improvements [46]. To ensure usability, evaluation should, ideally, be performed during development [14]. Conventional UEMs for user interfaces such as user surveys, observation and testing, and heuristic evaluation can be applied to identify problems in e-learning applications [32]. However, Squires and Preece [47] recommend that these approaches should be used differently in evaluating e-learning. Considering the discussion in Section 2.2, evaluation of educational software should investigate usability, interaction design, pedagogical effectiveness, learning content, and how well learners are supported in learning. Similarly, Masemola and De Villiers [31] point out that evaluating e-learning is different from evaluating conventional task-based software since, in the former context, the focus is on the *process* supported by the application (learning) rather than on a *product* generated by

interacting with the system. There should be an integration of usability, didactic effectiveness and learning issues in such evaluations [3, 8, 41, 47]. This is the approach taken in this study.

The selection of appropriate UEMs depends on various factors. Since some instructors and developers are unfamiliar with the methods [54] and cannot undertake evaluations themselves, it is important that inexpensive, effective and non-complex methods be used to evaluate usability and to determine usability problems [3].

#### 2.3.1 Heuristic evaluation

Heuristic evaluation (HE) is a usability inspection technique developed by the champion of usability, Jakob Nielsen [33, 34]. By definition of Dix, Finlay, Abowd and Beale [10, p. 320, 327], HE is classified as 'evaluation through expert analysis', which is distinguished from the category 'evaluation through user participation', i.e. it is inherent to HE that it is not conducted by actual users. During the process, a few expert evaluators conduct pre-defined, representative tasks on the application and, guided by a given set of usability principles known as heuristics (e.g. the criteria in Section 4), independently determine whether the interaction conforms to these principles [5, 15, 32, 39]. HE is a classic and widely-used UEM in frequent use. Researchers are using it in various contexts and variants, a number of them in evaluating educational websites [19, 30]. Nielsen describes it as 'discount usability engineering' [40, p. 718] in that it is fast and inexpensive, because it is conducted by evaluators who are acknowledged domain experts. In a well-balanced set of evaluators, some are experts in usability and others are experts in the content of the system being studied, i.e. subject-matter experts. Evaluators who are experts both in the domain area and in HCI are termed 'double experts' [24, p. 97]. The heuristics or 'rules of thumb' guide their critique of the design under evaluation. The result of HE is a list of usability problems in the system, according to the heuristics used or other issues the evaluators identify [10, 33].

Factors involved in selecting and inviting a balanced set of experts, are the number to use and their respective backgrounds. It is seldom possible for a single evaluator to identify all the usability problems. However, different evaluators or experts find different problems, which may not be mutually exclusive. Thus, when more experts are involved with an evaluation, more problems are discovered. Nielsen's [33] cost-benefit analysis demonstrated optimal value, with three to five evaluators identifying 65-75% of the usability problems. Despite this, the debate continues. Eleven experts were used in a study [4] to assess the usability of a university web portal. Law and Hvannberg [27] reject the 'magic five' and used eleven participants to define 80% of the detectable usability problems. In line with Nielsen [33], Karoulis and Pombortsis [24] determined that two to three evaluators who are 'double experts' will point out the same number of usability problems as three to five 'single experts'.

According to Nielsen [33], the evaluation process comprises: identification of heuristics, selection of evaluators, briefing of evaluators, the actual heuristic evaluation, and finally, aggregation of the problems. Sometimes severity rating, i.e. assigning relative severities to individual problems, can be performed to determine each problem's level of seriousness, estimated on a 3- or 5-point Likert scale. The experts can do ratings either during their HEs or later, after all the problems have been aggregated. The latter approach is advantageous [1, 28] since evaluators have the opportunity to consider and to rate

problems they did not identify themselves. This is the approach adopted in this study.

HE is the most widely-used UEM for computer system interfaces, since it is inexpensive, relatively easy and fast to perform, and can result in major improvements [5, 24]. It can be used early in design [10], but also on operational systems, as in this study. Despite these advantages, questions arise regarding its effectiveness in identifying user problems and the nature of problems it identifies [4, 17, 24, 33]. This study provides some answers to this debate.

There are advantages to using HE in tandem with user-based methods. Several recent studies advocate combining it with user testing (UT) and user surveys. In an interview with Preece and her colleagues [40], Jakob Nielsen terms this combination a ‘sandwich model’ when used in a layered style. Tan, Liu and Bishu [50] combined HE of e-commerce sites with UT (observation and interviews) and found the two methods complementary in addressing different kinds of problems. HE identified more problems than UT. Requirements for their nine expert evaluators were postgraduate courses in HCI and human factors of web design, and participation in at least one HE. In Thyvalikakath, Monaco, Thambuganipalle and Schleyer [51] comparative study of HE and UT on four computer-based dental patient records systems (CPRs), HE predicted on average 50% of the problems found empirically by UT. The UT involved think-aloud by 20 novice users, coded in detail by researchers. The three expert evaluators were dentists: two were postgraduate Informatics students who had completed HCI courses; the third was an Informatics faculty member and an expert in HE. They were familiar with CPRs in general, but had not used them routinely. This would qualify them as single experts. Hvannberg, Law and Lárusdóttir [18] combined the HE of an educational web portal with a set of user-based methods (observation, recording, questionnaires). As experts, they used 19 final-year BSc Computer Science students and one BSc (CS) graduate. The researchers acknowledge that these evaluators had a sound knowledge of evaluation but little practice. HE identified more problems than UT and 38% of the experts’ problems were confirmed by the user study. Also in an e-learning context, Ardito and his colleagues [3] posit that reliable evaluation can be achieved by systematic combination of inspection and user-based methods. In their study of Computer Science students learning HCI via the Internet, the user-based methods employed were think-aloud and interviews.

### 2.3.2 User-based surveys: Questionnaires and interviews

Some researchers believe that asking users is one of the best ways to identify usability problems [2, 10], the main survey (query) techniques being questionnaires and interviews. In evaluating e-learning, surveys can probe learner-centred issues that are not always obvious to experts in HE’s, for example, how easy it is for students to learn with a particular educational application.

Questionnaire surveys are established techniques of collecting demographic data and users’ opinions. They generally consist of closed or open question structures [40], where open questions allow participants to express spontaneous answers, whereas closed questions offer a set of options as answers. Though open questions provide rich data, they are more difficult to analyse than closed questions. Before carrying out a major survey, questionnaires should be prepared, reviewed, and pilot-tested with a small sample of users to avoid potential misunderstanding and to identify unfair questions, where respondents understand the questions, but lack the background to respond [12, 46]. Questions are predetermined

and fixed for all users. Questionnaires can reach a wide group of participants and are inexpensive and relatively simple to administer [10, 46]. However, they are not customised to individuals and less flexible than interviews.

In order to investigate further, questionnaire surveys are frequently accompanied by interviews, where qualitative information is gathered from individuals, by querying them verbally about the usability of the system. This can be very productive, since the interviewer can pursue unanticipated avenues and specific issues that may lead to focussed and constructive suggestions from the respondents [10, 46]. The level of questioning can be varied to suit the context and the interviewer can probe more deeply where appropriate [10]. After individual interviews, focus-group discussions or group interviews can be conducted to further explore issues arising [46].

## 3. RESEARCH DESIGN AND METHODOLOGY

### 3.1 Research question

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

*How effective is heuristic evaluation by experts in identifying usability- and learning-related problems in a web-based learning application?*

### 3.2 Design and methodology

The research design of this paper involves a study using two UEMs. We report on the methods and findings of HE by high-calibre expert evaluators, combined with user surveys in a comprehensive usability evaluation of a web-based learning application, considering both usability and pedagogical aspects. The study aims primarily to investigate the use of HE in the context of WBL and, to a lesser extent, to evaluate the target system. Various studies [11, 15, 18, 22, 50, 51] have been conducted to determine the relative effectiveness of different UEMs. They focused mainly on conventional software interfaces, unlike this study where the target system was a WBL environment for a specific higher-education course. We are aware of the comparative usability evaluation (CUE) debate – termed the ‘Damaged Merchandise’ controversy – regarding the meta-evaluation of UEMs themselves [15, 16, 36]. The debate is not addressed in this paper but, in a subsequent study, its stances should be applied to the approaches in the present study. This study first involved a form of *development research* in the generation of an evaluation framework comprising criteria structured into categories (Section 4). This framework was then applied in a *comparative usability evaluation* study, supplementing HE by experts with a user-based questionnaire survey. The results of the HE were analysed and compared with survey results to answer the research question. Sub-methods such as interviews and observations were also used to collect multiple-triangulated evidence. The yardstick used in determining the effectiveness of HE was the number and nature of problems identified by this method and how they related to those from the user-based evaluations. The main research processes were:

1. *Determination of criteria / heuristics*: Identification – based on the literature – of sets of criteria in appropriate categories for evaluating WBL applications, and their synthesis into an integrated evaluation framework.
2. *Heuristic evaluation*: Heuristic evaluation with a competent and complementary set of experts, followed by severity rating on the consolidated set of problems.

3. *User-based methods:* In Section 2 studies are mentioned that combine HE with user testing (UT). In the present study, formal UT was not feasible due to lack of financial resources for sophisticated usability-testing equipment as well as insufficient human resources for time-intensive manual observation. However, user-based methods in the form of criterion-based surveys were a satisfactory technique, with high reliability due to using almost the entire learner population as participants, followed by a focus group interview.
4. *Results and comparison:* Analysis of HE results and comparison of these with user-based survey results to answer the research question.  
These processes are covered in Sections 4 to 6.

### 3.3 The target application and its users

The evaluations were conducted on *Info3Net*, a custom-built course website for Information Systems (IS) 3 at Walter Sisulu University (WSU) in East London in the Eastern Cape. Among other topics, IS 3 covers *Advanced Databases*, for which *Info3Net* was designed and built using as development environment *WebCT* (now *Blackboard*), which is a sophisticated tool and course management system [41]. *Info3Net* was selected as the target for evaluation due to its convenience, accessibility and ease of control. It was designed by the first author (lecturer and subject co-ordinator of IS 3 at the time).

WBL environments can supplement face-to-face teaching and learning, or serve as stand-alone tools for instruction [21]. In this case, the former approach, called blended learning, was used. About 80 students in two streams, the final years of *Information Technology* and *Financial Information Systems*, respectively, used *Info3Net* to supplement traditional contact sessions. They were all competent in using computers.

## 4. CRITERIA FOR EVALUATING WEB-BASED LEARNING ENVIRONMENTS

As noted in Sections 1 and 2, evaluation criteria/heuristics (terms used interchangeably) for e-learning should address interfaces and usability principles from HCI as well as pedagogy and learning-related issues. Appropriate literature studies were undertaken, resulting in the synthesis of an integrated framework [48] of twenty criteria within three categories for evaluating WBL environments. Table 1, in the form of Tables 1a, 1b and 1c, presents heuristics for the three categories of *General interface criteria, customised for e-learning* (General); *Website-specific criteria* (Web); and *Educational criteria for learner-centred instructional design* (Educational), respectively. The 20 primary criteria are listed, along with their sub-criteria. Nielsen’s heuristics [33] form the basis of Category 1, with extensions influenced by Squires’ and Preece’s [47] ‘learning with software’ heuristics. Various other sources [1, 27, 28, 31, 43] were also used in the generation

process. The sub-criteria or guidelines for each criterion are customisable to the evaluation context. The expert evaluations excluded Criteria 12, 15, 16, 17 and 19, which focus on personal learning experiences of students, but are less relevant to expert evaluators. Consequently, 15 of the 20 criteria were common to both the expert and survey evaluations.

## 5. EVALUATION

The evaluations were conducted in a professional, ethical and socially-responsible manner. Participants’ rights were protected. Both the end-user participants and expert evaluators signed a consent form and received information documents, explaining the purpose, assuring anonymity and stating that findings would be used for research purposes only [40, 46]. Participation was voluntary and participants could withdraw at any time. Students received no reward for participating. Expert evaluators were not paid for their services, but received a gift as token of gratitude.

The next two subsections describe the approaches and methods of these evaluations.

### 5.1 Heuristic evaluation by experts

Our HE approach [49] was based on that advocated and applied by Nielsen [33], followed by severity rating. As custom-built heuristics relating to both usability- and learning-related guidelines, the synthesized set of criteria in Table 1 was used.

#### 5.1.1 Selection of evaluators

Debate exists on the optimal number and background of experts to participate in an HE (Section 2.3.1). This study adopted Albion’s approach [1] involving four experts. Four expert evaluators, jointly representing long-term expertise in user interface design, instructional/educational design, teaching, and subject matter, were invited and agreed to participate. Two were lecturers in database subject-matter domain as well as HCI experts, familiar with HE and can be classified as ‘double experts’. Table 2 shows the profiles of the four experts; all of them were academics who taught, did postgraduate supervision, and conducted research. They came from three different tertiary institutions in different cities.

#### 5.1.2 Briefing the evaluators

Evaluators were briefed in advance about the HE process for the study, the subject-matter domain, the target system, and the task scenarios to work through as advocated by [28] and [33]. In addition to the consent form and a request to familiarise themselves with the heuristics upfront, evaluators were given documents detailing the phases and processes up to severity rating. The system- and user profile documents gave a general background to *Info3Net* and explained how learners used it. The HE Procedure set out log-in details, scenarios to perform, instructions on how to do the actual evaluation, and how to compile the report of usability problems.

**Table 1. A comprehensive set of heuristics/criteria for evaluation of WBL environments [48]**

<b>Table 1 (a): Category 1: General interface usability criteria (based on Nielsen’s heuristics, modified for e-learning context)</b>	
1	<b>Visibility of system status</b> <ul style="list-style-type: none"> <li>• The website keeps the user informed through constructive, appropriate and timely feedback.</li> <li>• The system responds to user-initiated actions. There are no surprise actions by the site or tedious data entry sequences.</li> </ul>
2	<b>Match between the system and the real world i.e. match between designer model and user model</b> <ul style="list-style-type: none"> <li>• Language usage in terms of phrases, symbols, and concepts is similar to that of users in their day-to-day environment.</li> <li>• Metaphor usage corresponds to real-world objects/concepts, e.g. understandable and meaningful symbolic</li> </ul>

	<p>representations are used to ensure that the symbols, icons and names are intuitive within the context of the task performed.</p> <ul style="list-style-type: none"> <li>Information is arranged in a natural and logical order.</li> </ul>
3	<p><b>Learner control and freedom</b></p> <ul style="list-style-type: none"> <li>Users control the system.</li> <li>Users can exit the system at any time, even when they have made mistakes.</li> <li>There are facilities for Undo and Redo.</li> </ul>
4	<p><b>Consistency and adherence to standards</b></p> <ul style="list-style-type: none"> <li>The same concepts, words, symbols, situations, or actions refer to the same thing.</li> <li>Common platform standards are followed.</li> </ul>
5	<p><b>Error prevention, in particular, prevention of peripheral usability-related errors</b></p> <ul style="list-style-type: none"> <li>The system is designed in such a way that the users cannot easily make serious errors.</li> <li>When a user makes an error, the application gives an appropriate error message.</li> </ul>
6	<p><b>Recognition rather than recall</b></p> <ul style="list-style-type: none"> <li>Objects to be manipulated, options for selection, and actions to be taken are visible.</li> <li>The user does not need to recall information from one part of a dialogue to another.</li> <li>Instructions on how to use the system are visible or easily retrievable whenever appropriate.</li> <li>Displays are simple and multiple page displays are minimised.</li> </ul>
7	<p><b>Flexibility and efficiency of use</b></p> <ul style="list-style-type: none"> <li>The site caters for different levels of users, from novice to expert.</li> <li>Shortcuts or accelerators, unseen by novice users, are provided to speed up interaction and task completion by frequent users.</li> <li>The system is flexible to enable users to adjust settings to suit themselves, i.e. to customise the system.</li> </ul>
8	<p><b>Aesthetics and minimalism in design</b></p> <ul style="list-style-type: none"> <li>Site dialogues do not contain irrelevant or rarely needed information, which could distract users.</li> </ul>
9	<p><b>Recognition, diagnosis, and recovery from errors</b></p> <ul style="list-style-type: none"> <li>Error messages are expressed in plain language.</li> <li>Error messages define problems precisely and give quick, simple, constructive, specific instructions for recovery.</li> <li>If a typed command results in an error, users need not retype the entire command, but only the faulty part.</li> </ul>
10	<p><b>Help and documentation</b></p> <ul style="list-style-type: none"> <li>The site has a help facility and other documentation to support users' needs.</li> <li>Information in these facilities is easy to search, task-focused, and lists concrete steps to accomplish a task.</li> </ul>

*Table 1 (b): Category 2: Website-specific criteria*

11	<p><b>Simplicity of site navigation, organisation and structure</b></p> <ul style="list-style-type: none"> <li>The site has a simple navigational structure.</li> <li>Users should know where they are and have the option to select where to go next, e.g. via a site map or breadcrumbs.</li> <li>The navigational options are limited, so as not to overwhelm the user.</li> <li>Related information is placed together.</li> <li>Information is organised hierarchically, moving from the general to the specific.</li> <li>Common browser standards are followed.</li> <li>Each page has the required navigation buttons or hyperlinks (links), such as previous (back) next and home.</li> </ul>
12	<p><b>Relevance of site content to the learner and the learning process</b></p> <ul style="list-style-type: none"> <li>Content is engaging, relevant, appropriate and clear to learners using the WBL site.</li> <li>The material has no biases such as racial and gender biases, which may be deemed offensive.</li> <li>It is clear which materials are copyrighted and which are not.</li> <li>The authors of the content are of reputable authority.</li> </ul>

*Table 1 (c): Category 3: Educational criteria: Learner-centred instructional design, grounded in learning theory*

13	<p><b>Clarity of goals, objectives and outcomes</b></p> <ul style="list-style-type: none"> <li>There are clear goals, objectives and outcomes for learning encounters.</li> <li>The reason for inclusion of each page or document on the site is clear.</li> </ul>
14	<p><b>Effectiveness of collaborative learning (where such is available)</b></p> <ul style="list-style-type: none"> <li>Facilities and activities are available that encourage learner-learner and learner-teacher interactions.</li> <li>Facilities are provided for both asynchronous and synchronous communication, such as e-mail, discussion forums and chat rooms.</li> </ul>

15	<p><b>Level of learner control</b></p> <ul style="list-style-type: none"> <li>• Apart from controlling the interactions with the site, learners have some freedom to direct their learning, either individually or collaboratively, and to have a sense of ownership of it.</li> <li>• Learners are given some control of the content they learn, how it is learned, and the sequence of units.</li> <li>• Individual learners can customise the site to suit their personal learning strategies.</li> <li>• Educators can customise learning artefacts to the individual learner, for example, tests and performance evaluations can be customised to the learner’s ability.</li> <li>• Where appropriate, learners take the initiative regarding the methods, time, place, content, and sequence of learning.</li> </ul>
16	<p><b>Support for personally significant approaches to learning</b></p> <ul style="list-style-type: none"> <li>• There are multiple representations and varying views of learning artefacts and tasks.</li> <li>• The site supports different strategies for learning and indicates clearly which styles it supports.</li> <li>• The site is used in combination with other mediums of instruction to support learning.</li> <li>• Metacognition (the ability of a learner to plan, monitor and evaluate his/her own cognitive skills) is encouraged.</li> <li>• Learning activities are scaffolded by learner support and by optional additional information.</li> </ul>
17	<p><b>Cognitive error recognition, diagnosis and recovery</b></p> <ul style="list-style-type: none"> <li>• Cognitive conflict, bridging and problem-based learning strategies are used in the recognition-diagnosis-recovery cycle.</li> <li>• Learners have access to a rich and complex environment in which they can explore different solutions to problems.</li> <li>• Learners are permitted to learn by their mistakes and are provided with help to recover from cognitive errors.</li> </ul>
18	<p><b>Feedback, guidance and assessment</b></p> <ul style="list-style-type: none"> <li>• 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.</li> <li>• Learners are guided as they perform tasks.</li> <li>• Quantitative feedback, e.g. grading of learners’ activities, is given, so that learners are aware of their level of performance.</li> </ul>
19	<p><b>Context meaningful to domain and learner</b></p> <ul style="list-style-type: none"> <li>• Knowledge is presented within a meaningful and authentic context that supports effective learning.</li> <li>• Authentic, contextualised tasks are undertaken rather than abstract instruction.</li> <li>• The application enables context- and content-dependent knowledge construction.</li> <li>• Learning occurs in a context of use so that knowledge and skills are transferable to similar contexts.??</li> <li>• The representations are understandable and meaningful, ensuring that symbols, icons and names used are intuitive within the context of the learning task.</li> </ul>
20	<p><b>Learner motivation, creativity and active learning</b></p> <ul style="list-style-type: none"> <li>• The site has content and interactive features that attract, motivate and retain learners, and that promote creativity, e.g. the online activities are situated in real-world practice, and interest and engage the learners.??</li> <li>• To promote active learning and critical thinking, tasks require learners to compare, analyse and classify information, and to make deductions.</li> </ul>

**Table 2. Profiles of expert evaluators**

Evaluators	Evaluator 1	Evaluator 2	Evaluator 3	Evaluator 4
<b>Highest qualifications</b>	MEd; PhD	MCom (IS)	MEd; MSc (IS); PhD (IS)	MEd; DSc (Ed)
<b>Professional role</b>	Prof. in Dept. of Education.	Senior lecturer in IS. Course manager for IS Honours.	Prof. in IS Dept; Course manager for IS Masters.	Senior lecturer in Dept. of IT.
<b>Duties/courses taught (relevant to this study)</b>	ICT, Computer Literacy, Psychology of Education.	HCI at Honours level, Teaches Database Design at 3 <sup>rd</sup> level (same as <i>Info3Net</i> ).	HCI at Honours level, Teaches Database Design at 3 <sup>rd</sup> level (same as <i>Info3Net</i> ).	Research Skills for BTech, Internet Programming at 2 <sup>nd</sup> level.

5.1.3 Actual heuristic evaluation

Each expert evaluator conducted his/her evaluation independently. They performed the same tasks as those in the end-user surveys, applying the 15 selected criteria along with their sub-criteria. Over and above noting issues arising out of those tasks, experts were requested to assess general features of the application, just as the learners did. 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 achieved this. In discussions with the others, we agreed that the researcher would be available as an objective observer-facilitator during the evaluations, not to intervene, but to expedite the process by answering queries and serving as a scribe. This is recommended by [28] and [33], since it enables the evaluator to concentrate on pointing out usability problems. With this support, the other three experts performed their evaluations within the next ten days, taking less time than the first evaluator.

5.1.4 Severity rating of the problems

Once both expert- and end-user evaluations were complete, severity rating was conducted on the consolidated set of usability problems. The aggregated list of problems identified by the four expert evaluators was merged with those pinpointed by the learners to make a single list of all the identified problems, categorised according to the criteria they violated. This final consolidated set was sent to the expert evaluators for severity rating. To provide more information, each problem was accompanied by a weight, 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 completed the form within three days.

Table 3 shows the 5-point rating scale [38] used to assess the problems and assign severities. The scale is similar to Nielson’s [33] with an additional rating, ‘Medium’, between ‘Major’ and ‘Minor’.

**Table 3: Five-point rating scale for severity of usability problems [38]**

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

5.2 Survey evaluations among end-users (learners)

The design of the questionnaire [49] included closed and open questions. The main section was based on the 20 criteria in Table 1. For each criterion, straightforward, descriptor statements, based on the subcriteria, were generated to expand the meaning, since students might not understand the technical terms in the framework. This simplification of subcriteria to single-issue statements is shown by the extract in Table 4 relating to Criterion 3 of Category 1. Students rated *Info3Net* in this way, which provided valuable data, but the most important aim was for them to use the open-ended response areas provided with each criterion to explain problems they had experienced.

**Table 4. Extract from the Criterion 3 section of the questionnaire**

<b>3. User control and freedom</b>				
3.1 I control the system, rather than it controlling me.				
Strongly agree	Agree	Maybe	Disagree	Strongly disagree
3.2 The system works the way I want it to work.				

Strongly agree	Agree	Maybe	Disagree	Strongly disagree
3.3 Each page has all the required navigation buttons or hyperlink (link), such as Previous (Back), next (Forward) and Home.				
Strongly agree	Agree	Maybe	Disagree	Strongly disagree
3.4 When I make a mistake I can choose to exit (close) the system, using a clearly marked emergency Exit button.				
Strongly agree	Agree	Maybe	Disagree	Strongly disagree
Describe any particular problem(s) you encountered in <i>Info3Net</i> in relation to this section.				
<i>(Open-ended response area reduced in this article)</i>				

The questionnaire was pilot tested by five students. Feedback from the pilot and observation by a researcher identified minor problems in the instrument, which were rectified before the survey.

Sixty one of eighty students using *Info3Net* (76% of the population) participated in the survey, which was conducted on a single day in the students’ usual computer laboratory setting, in two separate groups. The questionnaire was followed by a focus group interview with eight students, as advocated by [46]. This clarified and elaborated problems and further problems were identified.

6. RESULTS

6.1 Heuristic evaluation results

6.1.1 Problems identified by expert evaluators

In total, 77 unique usability problems were initially identified by the set of four experts, showing that HE can play a meaningful role. To eliminate duplicates, the researcher carefully considered them and combined those that were closely related, resulting in a consolidated list of 58 problems. Table 5 shows the number of problems identified by each evaluator for each category, *General*, *Web*, and *Educational*. The *Consolidated* column gives the number of problems, after the consolidation process, with respect to a specific Heuristic. *By single expert* represents the number of problems identified by one expert only (not necessarily the same expert). For example, of the six problems identified by the set of experts against Heuristic 2, five were identified by single experts, meaning that only one was identified by more than one expert.

Table 5 shows that between 9 and 31 problems (16 - 53% of total) were recorded by each expert, with an average of 21. The fact that 41 of the 58 were identified by single experts, shows the value of a complementary team. Evaluator 1 identified less than the others. This is the only evaluator who performed it in the required period and without support of a researcher to clarify issues and serve as scribe. S/he did it in two hours, whereas the others, working with a researcher who recorded

problems while they evaluated, completed it in one to one-and-a-half hours, confirming suggestions by Nielsen and others [28, 33] that this support expedites the process.

The number and nature of problems identified by evaluators is associated with their backgrounds and expertise (Table 2). The

lower proportion of problems identified by the ‘single expert’, Evaluator 1, is attributed to her status as an expert in educational theory and educational technology. S/he had not

**Table 5: Number of problems identified by experts**

Category	Heuristic No.	Evaluator 1	Evaluator 2	Evaluator 3	Evaluator 4	Consolidated	By single expert
General	1	0	1	2	1	4	4
	2	0	2	4	1	6	5
	3	1	2	2	3	5	3
	4	0	1	2	2	3	2
	5	0	0	2	0	2	2
	6	1	3	4	1	5	3
	7	1	1	3	5	6	3
	8	1	1	1	0	3	3
	9	0	0	1	1	1	0
	10	1	1	1	0	3	3
	<b>Sub-Total</b>	<b>5</b>	<b>12</b>	<b>22</b>	<b>14</b>	<b>38 (66%)</b>	
Web	11	0	5	3	2	7 (12%)	4
Educational	13	0	0	2	2	4	4
	14	0	0	1	1	1	0
	18	4	2	2	2	6	4
	20	0	1	1	1	2	1
	<b>Sub-Total</b>	<b>4</b>	<b>3</b>	<b>6</b>	<b>6</b>	<b>13 (22%)</b>	
<b>Total</b>		<b>9</b>	<b>20</b>	<b>31</b>	<b>22</b>	<b>58</b>	<b>41</b>
<b>%</b>		<b>16</b>	<b>34</b>	<b>53</b>	<b>38</b>	<b>100</b>	<b>71</b>

studied HCI nor had s/he previously done an HE. The two ‘double experts’ [24], Evaluators 2 and 3, who identified higher proportions of problems, have experience in HCI and in the database subject matter of *Info3Net*. They both lecture postgraduate HCI. Evaluator 2 has particular expertise in heuristic evaluation and is an experienced educator, although without formal education training. Evaluator 3 has postgraduate qualifications in both IS and Education. S/he can be described as an ideal ‘double expert’ and identified the highest number of problems in the application.

Evaluator 4 was initially viewed as a single expert, yet s/he identified many problems. This could be because s/he designed e-learning applications using the same course management system used for designing *Info3Net*. Furthermore, this evaluator

lectures Internet programming, making him/her almost another ‘double expert’, whose experience as a designer and receiver of feedback from his/her own students enhanced recognition of problematic issues.

These results demonstrate that double experts with both domain and HCI knowledge discover high proportions of usability problems in line with Nielsen and Phillips [35], who found that usability specialists are better than non-specialists at finding usability problems. Ideally, HE usability experts should also have skills and experience for domain-specific evaluations [23]. Although HEs are valuable and authoritative ways of identifying usability problems, the findings depend on evaluators’ skills and experience and the contributions of different experts may vary.

The *By single expert* column in Table 5, represents the problems identified by one expert evaluator only, which constitute 71% (41 out of 58) of the experts' problems. This means that only 29% (100% minus 71%) of the problems were identified by more than one evaluator, which indicates a low level of agreement among the experts. Further analysis of Table 5 shows a tendency for each evaluator to identify several problems associated with a specific heuristic. For example, of the nine problems identified by Evaluator 1, four violated Heuristic 18. Likewise, five of the twenty identified by Evaluator 2 violated Heuristic 11, and five of Evaluator 4's twenty two problems related to Heuristic 7. This highlights the importance of using a varied group of evaluators to investigation an application from multiple perspectives.

Table 5 shows that the experts identified problems in the General, Web and Educational categories, except for Evaluator 1 who did not identify any in the Web category. When combined, the General category had the highest number of problems, 66% (38 problems out of 58) followed by the Educational category with 22% and Web with 12%. With hindsight, one notes that there are 10, 4 and 1 criteria in these categories respectively, and that the number of problems per category is proportional to the number of criteria. A possible conclusion is that a minimum number of criteria per category is required to support optimal identification of problems.

### 6.1.2 Solving the problems

Although HE focuses primarily on identifying problems, not solving problems, in most cases the solution is implicit in the problem. Experts identified problems in *Info3Net* that, if fixed, would improve it from both learning and usability perspectives. For example, Table 6 shows the problems that violate Heuristic 3. They are specific and most of the solutions can be determined directly from the associated problem statements.

**Table 6: Sample of problems identified by experts**

<b>Heuristic 3: User control and freedom</b>
<ul style="list-style-type: none"> <li>It is not easy to print site content, such as the Learner Guide and Content Modules. For example, there is no 'Print version' of the notes found in the hyperlinks in the Table of Contents.</li> </ul>
<ul style="list-style-type: none"> <li>There are no facilities for Undo and Redo.</li> </ul>
<ul style="list-style-type: none"> <li>Sometimes the system is slow to respond.</li> </ul>
<ul style="list-style-type: none"> <li>There is no system exit button.</li> </ul>
<ul style="list-style-type: none"> <li>There is no way to exit &lt;Help&gt; and return to the main system, apart from closing the Help window.</li> </ul>

### 6.2 User survey results

Sixty four different problems were identified by the 61 learners who participated in the survey. As in the case of HE, related problems were combined, resulting in 55 unique problems. The numbers of problems are shown in the *Students* column of Table 7, alongside the *Experts* column.

## 6.3 Comparison of results and discussion of findings

### 6.3.1 Overall comparison

Table 7 presents the overall results of the study. When the experts' and learners' problems were consolidated, there were 75 in all, of which 58 were identified by experts and 55 by learners, while 38 are common to both groups. Thus 77% were identified by the expert group, 73% by the learners, and 51% by both groups. Although the number of learner participants, namely 61, was significantly more than the number of experts, namely 4, the former large group identified fewer problems than the latter very small group. Since the experts identified approximately three-quarters of the problems (77%), the findings are in line with Nielsen and Molich's view that experts are not perfect in identifying problems. The results also support the belief [3, 6] that reliable evaluation can be achieved by systematically combining inspection methods such as HE with user-based methods, such as surveys.

More than half, 51% (38 of 75), of all problems were identified by both groups. With regard to cross-overs, Table 7 shows that the students identified 38 of 58 (66%) of the experts' problems while the experts identified 38 of 55 (69%) of the students problems. There is a notable correspondence between the types of problems identified by each group. A correlation coefficient of 0.55 between the numbers of problems identified by each group for each of the 15 criteria, shows a high correspondence, i.e. where students identified many problems with respect to a given criterion, so did the experts, and vice versa.

It is stated in Section 2 that HE is inexpensive and relatively easy to perform, compared to other UEMs. Issues of cost and researcher-time were not formally quantified in this study. However, considerably less time and effort was spent by the facilitator-researcher on the heuristic evaluation than on the survey, taking into account preparation of materials, conducting evaluations, and analysing the data. For example, only four sets of results were processed in the HE, versus 61 sets in the learner survey, yet, due to their expertise, the experts identified more problems.

### 6.3.2 Category comparison

Table 8 shows the number of problems identified by both sets of evaluators in the different categories. The table shows that 65%, 13%, and 23% of the aggregated (combined) problems belong to the General, Web and Educational categories respectively. A quarter (23%) of the problems identified were learning-related and the rest (78%) were interface usability problems, justifying the use of a comprehensive set of criteria.

Table 8 also shows that both sets of evaluators identified problems across the three categories. Of the problems in *Experts* column, 66%, 12% and 22% were in the General, Web and Educational categories respectively (Section 6.1.1). This is notably similar to the 67%, 9% and 24% respectively in the *Students* column of the table. It indicates that the proportion of problems identified by both groups in the three categories match closely.

**Table 7: Number of problems identified by students and by experts [49]**

Category	Criterion	Number of problems identified			
		Experts	Students	Common	All
<b>Category 1: General interface design</b> (49 problems = 65 % of Total)	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
<b>Category 2: Educational website-specific aspects</b> (10 problems = 13% of Total)	11	7	5	2	10
<b>Category 3: Learner-centred instructional design</b> (16 problems= 23 % of Total)	13	4	3	3	4
	14	1	2	1	2
	18	6	6	4	8
	20	2	2	2	2
<b>Total (all problems)</b>		<b>58</b>	<b>55</b>	<b>38</b>	<b>75</b>
<b>% of all 75 problems</b>		77%	73%	51%	100%
<b>% of other group's problems</b>		69%	66%		
<b>Correlation coefficient</b>		0.55			

**Table 8: Number and proportion of problems identified by experts and students**

	Experts		Students		All (combined)	
	Number	%	Number	%	Number	%
<b>General</b>	38	<b>66</b>	37	<b>67</b>	49	<b>65</b>
<b>Web</b>	7	<b>12</b>	5	<b>9</b>	10	<b>13</b>
<b>Educational</b>	13	<b>22</b>	13	<b>24</b>	16	<b>23</b>
<b>Total</b>	58	100	55	100	75	

6.3.3 Comparison based on severity of problems

Section 5.1.4 explains that the four expert evaluators independently rated the severity of the problems in the combined list of 75, based on a rating scale of 1 to 5 for *cosmetic-* to *catastrophic* problems respectively (Table 3). The mean of the average ratings was 3.0 with a standard deviation of 0.8, indicating that most problems were viewed as *medium* problems, 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. Examination of the source data shows that most problems were allocated very similar scores by all the expert evaluators. Moreover, the experts were not biased, indicated by the fact that each assigned high ratings to certain problems they had not recognised themselves. In so doing, they acknowledged and 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 and a large number of end users.

The general comments and the mean score of 3.0 assigned to the problems indicate that the experts rated the site positively in terms of its usability and support for learning. This was confirmed by qualitative data in the open-ended response areas of the questionnaire [49]. Two examples of students' comments were: '*The site is well organised and easy to navigate*' and '*I like the newness in learning – different learning styles. You feel like studying – it inspires you*'.

6.3.4 Comparison with respect to major and minor problems

Problems can be major or minor. Following the severity rating, those with an average of 4 to 5 were classified as *major*

problems and those with an average of 1 to 2 as *minor* problems (Table 3). After this categorisation, further comparison was undertaken of the findings of experts and students. Table 9 shows the number of problems and their corresponding percentages in the major and minor categories. Twelve problems were classified as major and fourteen minor, while the problems with ratings between 2 and 4 were considered neither major nor minor. Table 9 shows that of the twelve major problems, eleven were identified by experts, eleven by learners, and ten by both groups, namely, 92%, 92% and 83%. Secondly, the fact that ten problems were common indicates that of the major problems, ten of the eleven identified by learners, 91%, were also identified by the experts, compared to 69% of the learners' problems that were identified by experts when ALL problems were considered (Table 7 and Section 6.3.1). This data, in particular the fact that both the expert- and learner evaluators identified 92% of major problems, demonstrates very high correspondence. Of the twelve problems in the major category, only one identified by experts but not by learners, and one by learners but not by experts.

When considering minor problems, no such consensus is evident. Table 9 shows that of the fourteen minor problems, the experts identified eight (64%), learners five (43%) and only one (7%) was common to both groups. Both sets of evaluators identified a low percentage of minor problems and there was little agreement. These results support similar studies [26, 33], which show that major problems are easier to find than minor ones. Nielsen [33] found that, although major usability problems are easier to identify, when all problems are considered, the number of minor problems is likely to exceed the number of major ones. This finding is supported in the present study.

**Table 9: Problems identified by severity rating as major and minor problems**

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

## 7. CONCLUSIONS

To conclude, we re-visit the research question:

*How effective is heuristic evaluation by experts in identifying usability- and learning-related problems in a web-based learning application?*

The question was addressed by an investigation that can be described as 'meta-evaluation' in that it focused primarily not on the target system, but rather on analysing the numbers and nature of problems identified in the heuristic evaluation and validating the HE by comparing the results with those of an end-user survey among the learner population. Both studies were based on evaluation criteria custom-generated for web-based e-learning applications and divided into three categories targeting general interface-, website-specific- and learning-

related issues, respectively. The next two subsections summarise the findings, each ending with *success factors* and briefly discussing this research in the context of the literature.

### 7.1 Numbers and nature of problems identified by heuristic evaluation

A high number of real problems, originally 77 and consolidated to 58, was identified by four expert evaluators using fifteen of the twenty criteria. Some were usability issues; others related to learning support. The problems came from all three categories, with the number of problems in each being proportional to the number of criteria in that category. These findings show that HE can indeed be effective in evaluating WBL applications.

All the experts performed the evaluations with diligence and skill, but the double experts identified far more problems than the single expert.

#### Success factors and discussion

- Within the fifteen criteria were many (41) subcriteria that supported the evaluators in broad reflective processes.
- The expert evaluators were of a high calibre, with varying expertise and experience. In a complementary approach, they concentrated on different criteria and identified different problems. Four experts identified 77% of the aggregated set of problems, slightly exceeding the figures in Nielsen's [33] recommendation that three to five evaluators should identify 65-75% of the problems
- Double experts brought synergistic skills to the task. This finding is in line with studies [23, 35] suggesting that evaluators with both domain and HCI knowledge are particularly effective and discover a high proportion of problems, especially if they are also experienced in evaluation.

### 7.2 Findings of heuristic evaluation compared with the user-survey

User-based studies are the standard against which the effectiveness of HE is judged. The results of heuristic evaluation by experts corresponded well with those of survey evaluation among the end-users (learners). In fact, the HE results are better than the user-based results (see Table 7). They were produced by only four experts compared to 61 learners. Moreover, the experts were having their first encounter with *Info3Net*, whereas the learners had used it for a semester. Despite this, the students identified a slightly lower percentage of the aggregated set of problems, 73%, than the 77% by the expert evaluators. When totals are considered, the experts found 58 problems compared to the 55 identified in the learner survey.

When major problems only were considered, the proportion of problems identified rose to 92% for both groups. Furthermore, considering major problems only, the percentage of learner-problems identified by experts was 91%, compared with 69% against the set of all problems. There was a positive statistical correlation between the problems identified by the two groups i.e. where students identified many problems under a given criterion, so did the experts, and vice versa.

#### Success factors and discussion

- The close correspondence could be due, first, to the insight and perceptiveness of the experts (Table 3). They were mature senior academics with broad experience totalling over 80 years of academia, research and practice. In combination, they identified most of the problems in *Info3Net*.

- A second factor is the involvement of 76% of the students from two cohorts as participants in the survey, rather than samples of students or sets of 'users' who are not users in the real world. The close correspondence between the findings of two different UEMS contrasts with some of the literature (Section 2.3.1, latter part), where greater differences were found between sets of problems identified by different methods.
- We acknowledge that some of the other studies applied user testing, which is highly effective in identifying users' problems. When user testing is well done, it strengthens the user-study. Although formal usability testing was not feasible in this study, a success factor in identifying users' problems was the inclusion of open-ended questions in each section of the questionnaire, where learners could spontaneously describe problems they had encountered (Table 4). There were also follow-up focus-group interviews. These approaches identified further problems so that, despite the absence of user testing, the user-based study was strong. It effectively addressed the users' issues and improved reliability of the findings.
- Severity rating was used to distinguish the major problems. In the light of this classification, there was a close correspondence between the two sets of results.

### 7.3 Further contributions

- *Secondary benefit:* The evaluation of the *Info3Net* application, in and of itself, held value. Although it was rated positively in general, problems were identified that can be addressed.
- *Evaluation instrument:* The evaluation framework of categories, criteria, and sub-criteria synthesised for this research, is a flexible instrument that can be applied to identify usability- and pedagogical problems in e-learning environments. It is customisable to other applications and to a variety of UEMs. Since its publication [48, 49], there have been international requests to use it in research and teaching.
- *Design aid:* The framework can be used as a design aid for e-learning applications. The intention is not that every application or interactive learning environment should conform to all criteria, but the guidelines in the framework can be contextualised and adapted as required.
- *Comparative usability evaluation debate:* Future research could be undertaken to apply the points on both sides of the CUE debate about evaluation of UEMs – the 'Damaged Merchandise' controversy (see Introduction) – to the approaches in this study.

Finally, which UEM? The study indicates that heuristic evaluation, conducted by a skilled, experienced, competent and complementary group of experts, addressing web-, usability- and learning-related problems, is a highly effective usability evaluation method in the context of web-based learning. It is cost-effective and relatively easy to conduct. The identification level of major problems was high. We recommend that HE should, ideally, be supplemented with user-based methods, such as surveys or controlled usability testing. In cases where one approach only must be selected, the findings of this research can be used to propose heuristic evaluation as the optimal method.

### ACKNOWLEDGEMENTS

Sincere appreciation to the expert evaluators from Rhodes University, University of Fort Hare and Walter Sisulu

University (WSU) for their valuable contributions. One of these was Prof Rinette Roets, who has sadly passed away. Gratitude also to the IS 3 students at WSU for their keen participation.

### 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., Plantamura, P., Roselli, T., Rossano, V. and Tersigni, M. 2004. Towards Guidelines for Usability of e-Learning Applications. In: C. Stary and C. Stephanidis. (Eds), *Lecture Notes in Computer Science. Vol 3196, 2004: User-Centered Interaction Paradigms for Universal Access in the Information Society: 8th ERCIM Workshop on User Interfaces for All 2004*: 185-202. Berlin: Springer-Verlag.
- [3] Ardito, C., Costabile, M.F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T. and Rossano, V. 2006. An Approach to Usability Evaluation of E-Learning Applications. *Universal Access in the Information Society*, 4(3): 270-283.
- [4] Avouris, N., Tselios, N., Fida, C. and Papachristos, E. 2003. Website Evaluation: A usability-Based Perspective. In: Y. Manolopoulos, S. Evripidou and A.C. Kakas. (Eds), *Lecture Notes in Computer Science. Vol 2563: Advances in Informatics, 8th Panhellenic Conference on Informatics*: 217-232. Berlin: Springer-Verlag.
- [5] Blandford, A., Keith, S., Connell, I. and 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] Costabile, M.F., De Marsico, M., Lanzilotti, R., Plantamura, V.L. and 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.
- [7] De Villiers, M.R. 2005. E-Learning Artefacts: Are They Based on Learning Theory? *Alternation* 12 (1b): 345-371.
- [8] De Villiers, M.R. 2006. Multi-Method Evaluations: Case Studies of an Interactive Tutorial and Practice System. In: *Proceedings of InSITE 2006 Conference*. Manchester, United Kingdom.
- [9] De Villiers, M.R. 2007. An Action Research Approach to the Design, Development and Evaluation of an Interactive E-Learning Tutorial in a Cognitive Domain. *Journal of Information Technology Education*, 6: 455-479.
- [10] Dix, A.J., Finlay, J.E., Abowd, G.D. and Beale, R. 2004. *Human-Computer Interaction*. 3rd Ed. Harlow Assex: Pearson Education Limited.
- [11] Doubleday, A., Ryan, M., Springett, M. and Sutcliffe, A. 1997. A Comparison of Usability Techniques for Evaluating Design. *Proceedings of the 2nd Conference on Designing Interactive Systems: Process, Practice, Methods, and techniques*: 101-110. Amsterdam: ACM International Conference Proceeding Series.
- [12] Gillham, B. 2000. *Developing a Questionnaire*. London: Bill Gillham.

- [13] Govindasamy, T. 2002. Successful Implementation of E-Learning Pedagogical Considerations. *The Internet and Higher Education*, 4(4): 287-299.
- [14] Granic, A., Glavinic, V. and Stankov, S. 2004. *Usability Evaluation Methods for Web-Based Educational Systems*. [http://www.ui4all.gr/workshop2004/files/ui4all\\_proceedings/adjunct/evaluation/28.pdf](http://www.ui4all.gr/workshop2004/files/ui4all_proceedings/adjunct/evaluation/28.pdf).
- [15] Gray, W.D. and Salzman, M.C. 1998. Damaged Merchandise? A Review of Experiments that Compare Usability Evaluation Methods. *Human-Computer Interaction*, 13 (3): 203-261.
- [16] Gray, W.D. and Salzman, M.C. 1998. Repairing Damaged Merchandise: A Rejoinder. *Human-Computer Interaction*, 13 (3): 325-335.
- [17] Hollingsed, T. and Novick, D.G. 2007. Usability Inspection Methods after 15 Years of Research and Practice. *Proceedings of the 25th annual ACM international conference on Design of communication*: 249 – 255. El Paso: ACM Press.
- [18] Hvannberg, E.T., Law, E. L-C. and Lárusdóttir, M.K. 2007. Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with Computers*, 19: 225-240.
- [19] Hwang, G.J., Huang, T.C.K. and Tseng, J.C.R. 2004. A group-decision approach for evaluating educational web sites. *Computers & Education*, 42: 65-86.
- [20] ISO. 1998. *ISO-9241: Guidance on Usability Standards*. <http://www.iso.ch/iso/en/CatalogueListPage.CatalogueList?ICS1=13&ICS2=180>.
- [21] Jackson, R.H. 2004. *Weblearning Resources*. <http://www.knowledgeability.biz/weblearning/default.htm>.
- [22] Jeffries, R., Miller, J.R., Wharton, C. and Uyeda, K.M. 1991. User Interface Evaluation in the Real World: A Comparison of Four Techniques. In: S.P. Robertson, G.M. Olson and J.S. Olson. (Eds), *Proceedings ACM CHI'91 Conference*: 119-124. New Orleans: ACM Press.
- [23] Kantner, L. and Rosenbaum, S. 1997. Usability Studies of WWW Sites: Heuristic Evaluation vs. Laboratory Testing. In: *Proceedings of the 15th Annual International Conference on Computer Documentation 1997*: 153-160. Salt Lake City: ACM Press.
- [24] Karoulis, A. and 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.
- [25] Khan. B.H. 1997. *Web-Based Instruction*. New Jersey: Educational Technology Publications.
- [26] Law, L.C. and 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.
- [27] Law, L.C. and Hvannberg, E.T. 2004. Analysis of Combinatorial User Effect in International Usability Tests. *Proceedings of CHI2004*: 9-16.
- [28] Levi, M.D. and Conrad, F.G. 1996. A Heuristic Evaluation of a World Wide Web Prototype. *Interactions*, 3(4): 50-61.
- [29] Li, A., Shih, T.K., Li, F.W.B. 2008. Technology Support for Distributed and Collaborative Learning over the Internet. *ACM Transactions on Internet Technology*, 8(2):10 – 24.
- [30] Ling, C. and Salvendy, G. 2009. Effect of evaluators' cognitive style on heuristic evaluation: Field dependent and field independent evaluators. *International journal of human-computer studies*, 67: 382-293.
- [31] Masemola, S.S. and 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 and D. Kourie. *Service-Oriented Software and Systems. Proceeding of SAICSIT 2006*: 187-197. ACM International Conference Proceedings Series.
- [32] Minocha, S. and Sharp, H. 2004. Learner-Centred and Evaluation of Web-Based E-Learning Environments. *The 7th HCI Educators Workshop: Effective Teaching and Training in HCI*. Preston united Kingdom.
- [33] Nielsen, J. 1994. Heuristic Evaluations. In: J. Nielsen and R.L. Mack. (Eds), *Usability Inspection Methods*. New York: John Wiley & Sons.
- [34] Nielsen, J. and 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.
- [35] Nielsen, J. and Phillips, V.L. 1993. Estimating the Relative Usability of Two Interfaces: Heuristics, Formal, and Empirical Methods Compared. In: *Proceedings of INTERCHI 1993*: 214-221. New York: ACM Press.
- [36] Olson, G.M. and Moran, T.P. 1998. Commentary on "Damage Merchandise?" *Human-Computer Interaction*, 13 (3): 263-323.
- [37] Parlangeli, O., Marchingiani, E. and Bagnara, S. 1999. Multimedia in Distance Education: Effects of Usability on Learning. *Interacting with Computers*, 12(1): 37-49.
- [38] Pierotti, D. 1996. *Usability Techniques: Heuristic Evaluation Activities*. <http://www.stcsig.org/usability/topics/articles/he-activities.html>.
- [39] Preece, J. 1993. *A Guide to Usability: Human Factors in Computing*. The Open University: Addison-Wesley.
- [40] Preece, J., Rogers, Y. and Sharp, H. 2007. *Interaction Design: Beyond Human-Computer Interaction*. 2nd Ed. New York: John Wiley & Sons.
- [41] Quintana, C., Carra, A., Krajcik, J. and Soloway, E. (2002). Learner-Centered Design: Reflections and New Directions. In: J.M. Carroll (Ed.), *Human-Computer Interaction in the New Millennium*. New York: Addison-Wesley.
- [42] Razaq L. and Heffernan N.T. 2008. Towards Designing a User-Adaptive Web-Based E-Learning System. *CHI '08 extended abstracts on Human factors in computing systems*: 3525-3530. Florence: ACM Press.
- [43] Reeves, T.C. and 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.
- [44] Romiszowski, A.J. (2004). How's the E-learning Baby? Factors Leading to Success or Failure of an Educational

- Technology Innovation. *Educational technology*, 44(1):5-27.
- [45] Sandars J. and Hawthornthwaite C. 2007. New horizons for e-learning in medical education: ecological and Web 2.0 perspectives. *Medical Teacher*, 29:307-310.
- [46] Shneiderman, B. and Plaisant, C. 2005. Designing the User Interface: Strategies for Effective Human-Computer Interaction. 4rd Ed. New York: Addison-Wesley.
- [47] Squires, D. and Preece, J. 1999. Predicting Quality in Educational Software: Evaluating for Learning, Usability and the Synergy Between them. *Interacting with Computers*, 11(5): 467-483.
- [48] Ssemugabi, S. and 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.
- [49] Ssemugabi, S. and De Villiers, M.R. 2007. A Comparative Study of Two Usability Evaluation Methods Using a Web-Based E-Learning Application. In: Barnard, L. and Botha R.A. *Riding the wave of technology. Proceeding of SAICSIT 2007*: 132-142. ACM International Conference Proceedings Series.
- [50] Tan, W-S., Liu, D. and Bishu, R. 2009. Web evaluation: Heuristic evaluation versus user testing. *International Journal of Industrial Ergonomics*, 39: 621-627.
- [51] Thyvalikakath, T.P., Monaco, V., Thambuganipalle, H. and Schleyer, T. 2009. Comparative study of heuristic evaluation and usability testing methods. *Study of Health Technology Information*, 143: 322-327.
- [52] Van Greunen, D. And Wesson, J.L. 2005. Exploring Issues for information Architecture of Web-based Learning in South Africa. *South African Computer Journal*, 35: 112-117.
- [53] Veldof, J.R. 2003. Usability Tests. In: E.A. Dupuis. (Ed.), *Developing Web-Based Instruction: Planning, Designing, Managing, and Evaluating for Results*. London: Facet Publishing
- [54] Vrasidas, C. 2004. Issues of Pedagogy and Design in E-learning System. In: *ACM Symposium on Online Learning*: 911-915. Nicosia: ACM Press.
- [55] Wesson, J.L. and 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.
- [56] 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.
- [57] Zaharias, P. 2008. Developing Usability Evaluation Methods for E-Learning Applications: From Functional Usability to Motivation to Learn. *Journal of Human Computer Interaction*. In Press.