Campus-Wide Information Systems

Emerald Article: Heuristic evaluation of e-learning courses: a comparative analysis of two e-learning heuristic sets

Panagiotis Zaharias, Panayiotis Koutsabasis

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1065-0741.htm

RESEARCH PAPER

Heuristic evaluation of e-learning courses: a comparative analysis of two e-learning heuristic sets

Panagiotis Zaharias
Department of Information Systems, Open University of Cyprus, Nicosia, Cyprus, and
Panayiotis Koutsabasis
Department of Product and Systems Design Engineering, University of the Aegean, Hermoupolis, Greece

Abstract

Purpose – The purpose of this paper is to discuss heuristic evaluation as a method for evaluating e-learning courses and applications and more specifically to investigate the applicability and empirical use of two customized e-learning heuristic protocols.

Design/methodology/approach – Two representative e-learning heuristic protocols were chosen for the comparative analysis. These protocols augment the "traditional" heuristic sets so as to cover technology-enhanced learning properties. Two reviewers that have considerable experience in usability evaluation as well as in e-learning were involved in this comparative analysis. Coverage, distribution and redundancy were employed as three basic criteria for conducting the evaluation

Findings – The main results of the study indicate that both heuristic protocols exhibit wide coverage of potential usability problems. The distribution of usability problems is uneven to a large number of heuristics for both heuristic sets, which reveals that some heuristics are more general than others.

Originality/value – This study shows the empirical application of two heuristic protocols in a usability evaluation of e-learning applications. Furthermore, it provides a comparison of the two heuristic sets according to a set of criteria and provides a first set of suggestions regarding further development and validation of these heuristic sets.

Keywords Heuristic evaluation, Asynchronous e-learning courses, Technical and pedagogical usability, Usability problems, Function evaluation, E-learning, Problem solving

Paper type Research paper

1. Introduction

Heuristic evaluation (HE) is a systematic inspection of a user interface design for usability (Nielsen and Molich, 1990). The goal of heuristic evaluation is to find the usability problems in a user interface design so that they can be attended to as part of an iterative design process. It is the most commonly used inspection technique; it is inexpensive relative to other evaluation methods and intuitive and easy to motivate potential evaluators to use it while advanced planning is not required (Nielsen and Molich, 1990).

On the contrary heuristic evaluation does not provide a systematic way to generate solutions to the usability problems. It is a quite unstructured method and does not focus on user tasks (Ling and Salvendy, 2005). Additionally it does not involve real

Campus-Wide Information Systems
Vol. 29 No. 1, 2012
pp. 45-60

1065-0741
DOI 10.1108/10650741211192046

users, rather it is solely based on the evaluator's subjective judgment and thus, it has some degree of subjectivity (Jeffries *et al.*, 1991; Dumas and Redish, 1993). The most important drawback is that HE produces many false positive usability problems, which mean unreal usability problems that actual users may never face.

Some indicative problems of HE emerged from the relevant research agenda:

- Improving effectiveness and efficiency of heuristic evaluation with respect to user testing (Hvannberg *et al.*, 2006).
- Augmenting the "traditional" heuristic sets so as to cover new areas of computing such as home technology, learning technology and ubiquitous computing (Hornbaek, 2006; Ling and Salvendy, 2005).
- Identifying the appropriate tasks and reducing the selection bias (Cordes, 2001).
- Matching usability problem descriptions (this is an issue for any type of usability evaluation) (Hornbæk and Frøkjær, 2008) etc.

HE has been extensively used in the e-learning field (Albion, 1999; Reeves *et al.*, 2002; Ardito *et al.*, 2004; Pipan *et al.*, 2006; Zaharias, 2007). Nevertheless very little has been done to systematically confront the aforementioned research problems in e-learning context. Respective research efforts in e-learning studies are scarce. The effectiveness of the application of the method is questionable in many studies; very few efforts have been made to investigate such issues (effectiveness, efficiency etc) and most of the researchers and practitioners simply apply methods and heuristics from traditional usability research to study and evaluate e-learning environments. However a consensus matures in e-learning community, which demands an extended view of usability. Numerous researchers have realized that simple application of software heuristics could not effectively be applied because they fail to address the specific challenges of designing and evaluating modern interfaces for learners (Squires and Preece, 1999; Reeves *et al.*, 2002; Mehlenbacher *et al.*, 2005; Zaharias, 2007). Several augmented heuristic sets have been proposed but empirical validation is still lacking. This study discusses the heuristic evaluation method in e-learning context and focuses on the comparison of two customized to e-learning heuristic sets that have been proposed in the literature. It also presents and discusses the preliminary findings of the evaluation of an asynchronous e-learning course by employing the two different heuristic sets in parallel. This paper concludes by providing a further discussion of issues and research challenges for the heuristic evaluation method in the e-learning context.

2. Usability evaluation and the application of the heuristic evaluation method in e-learning studies

Usability evaluation of e-learning applications requires special attention due to the specificity and particularity of the learning process itself, the intricate characteristics of the learners and the variety of the learning activities. Accordingly the very nature of almost any e-learning application is quite complex. During the early days of the e-learning phenomenon usability was either ignored or applied simplistically, which mean that most of the evaluation efforts were conducted in "traditional" terms. However evaluation of usability in "traditional" terms is not enough (Zaharias, 2007; Zaharias and Poulymenakou, 2009). E-learning broadens the traditional definitions of

user, task and context derived from usability to account for formal and informal learning environments (Mehlenbacher *et al.*, 2005). It is widely accepted that e-learning environments are directly related to their pedagogical value (Ardito *et al.* 2006; Albion, 1999; Quinn, 1996, Squires and Preece, 1999; Silius and Tervakari, 2003). Evaluation of pedagogical usability should assess whether the pedagogical design of the learning environment is based on learning theories and whether other important factors such as motivation, diversity and growth (Soloway *et al.*, 1994) are taken into consideration (Silius and Tervakari, 2003). A body of researchers has started to work upon this direction during the recent years. Some researchers have developed models and criteria evaluation for learning management systems (Pipan *et al.*, 2006; Georgiakakis *et al.*, 2005). Others have focused on the usability of e-learning courses (Ardito *et al.*, 2004; Zaharias and Poulymenakou, 2009, Nokelainen, 2006, Dringus and Cohen, 2005). As the respective e-learning research matures, the common denominator is that usability of e-learning systems and applications is directed toward optimising the learner experience along the two main axes: technical (or web) usability and pedagogical usability.

Heuristic evaluation has been used extensively in e-learning studies (Albion, 1999; Reeves *et al.*, 2002; Zaharias, 2007). Many of the studies employ Nielsen's heuristics or slight variations of them without special consideration for e-learning specificities (Dringus, 1995b; Quinn, 1996; Albion, 1999). Albion (1999) adapted the heuristics proposed by Nielsen and Quinn (1996) and further proposed a set of heuristics customized for content of educational multimedia. Squires and Preece (1999) proposed an adaptation of Nielsen's (1994) heuristics, taking into account socio-constructivism tenets (Phillips, 1995; Soloway *et al.*, 1994) and the outcome was a set of eight "learning with software" heuristics:

(1) Match between designer and learner models.

(2) Navigational fidelity.

(3) Appropriate levels of learner control.

(4) Prevention of peripheral cognitive errors.

(5) Understandable and meaningful symbolic representations.

(6) Support personally significant approaches to learning.

(7) Strategies for cognitive error recognition, diagnosis and recovery.

(8) Match with the curriculum.

Most of these efforts can be characterized as initial attempts towards the integration of usability and learning and need further discussion, elaboration and validation. Another study conducted by Reeves *et al.* (2002) seems to provide a more elaborated tool for heuristic usability evaluation of e-learning programs. This work was based on Nielsen's protocol (1994) but expanded the original heuristics so as to include instructional design heuristics and thus providing a more appropriate method for the evaluation of e-learning applications. Another relevant study by Mehlenbacher *et al.* (2005) proposed a conceptual framework for the design and evaluation of all instructional situations. This is a holistic approach based on human-computer interaction (HCI) research and classical and contemporary rhetorical theory. The two aforementioned heuristic sets seem to follow a well-balanced integration of usability

and learning design and pedagogical issues which are necessary for the design and evaluation of modern e-learning applications; these two are the ones being compared and tested in this study.

In conclusion, according to the literature two streams of research can be delineated: The first is solely based on Nielsen's work and heuristics, which is the main broad stream. Research efforts in this stream try to extend the initial heuristic set in order to adequately cover usability issues in new computing areas such as e-learning. The main problem is that heuristic sets are mostly based on Nielsen's heuristics, which have received a lot of criticism about their validity. It is accepted that there is a lack of a strong theoretical background to explain how they work (Hvannberg *et al.*, 2006). In addition as already stressed they are quite generic and further elaboration is needed in specific fields of application such as e-learning. On the other hand, the second stream of research focuses on prior learning research, learning models and theories. Some of these models may be more appropriate in the e-learning context but most of the current relevant heuristic sets are in an initial state and need further improvement and empirical testing. For both of these streams, authors argue that for researchers who have proposed new sets (or for those who have proposed an expanded version of Nielsen' heuristics), the main problem is that they need further elaboration and empirical validation.

In this paper two customized e-learning heuristic sets, which represent in a great extent the main research streams respectively, are employed in an evaluation study of a typical asynchronous e-learning course. The main goal is to compare the two sets so as to identify their strengths and weaknesses and provide more empirical data towards their validation.

3. E-learning heuristic protocols: Reeves *et al.* vs Mehlenbacher *et al.*

As already mentioned this study focuses on the comparison of two different heuristic protocols that have been developed specifically for e-learning applications. The first one has been developed by Reeves *et al.* (2002) while the second one by Mehlenbacher *et al.* (2005). The following section briefly presents these two heuristic protocols.

3.1 The Reeves et al. *protocol*

This work was based on Nielsen's protocol (1994) but expanded the original heuristics so as to include instructional design heuristics and thus providing a more appropriate method for the evaluation of e-learning applications. The main contribution was 15 heuristics accompanied with a *Protocol for E-Learning Heuristic Evaluation* that explains how the whole process must be conducted. The main contribution was 15 heuristics accompanied with a *Protocol for E-Learning Heuristic Evaluation* that explains how the whole process must be conducted. This process contains eight steps and the usability problems are supposed to be evaluated along two scales. The 15 proposed heuristics emerged from the combination of Instructional Design and Usability heuristics:

(1) Visibility and System Status.

(2) Match between System and Real World.

(3) Error Recovery and Exiting.

(4) Consistency and Standards.

(5) Error Prevention.
(6) Navigation Support.
(7) Aesthetics.
(8) Help and Documentation.
(9) Interactivity.
(10) Message Design.
(11) Learning Design.
(12) Media Integration.
(13) Instructional Assessment.
(14) Resources.
(15) Feedback.

Several researchers have already highlighted the importance or used the Reeves *et al.* (2002) heuristics (Dringus and Cohen, 2005; Nguyen and Chang, 2006; Hartwig *et al.*, 2003; Ling and Salvendy, 2005; Zaharias, 2007). Nevertheless there is no empirical data regarding its effectiveness and efficiency in comparison with other proposed heuristics or techniques.

3.2 The Mehlenbacher et al. *protocol*

This work has been influenced by usability research that is grounded in both early Human-Computer Interaction (HCI) research and in classical and contemporary rhetorical theory as well as e-learning design.

This work is original in the sense that combines design, human-computer interaction research, rhetorical theory, and usability within the context of online instructional materials. By reviewing instruction, learning, and technology within the context of numerous research-based models of teaching and learning, Mehlenbacher *et al.* (2005) seek to identify common attributes of what they call "everyday instructional situations". In fact they propose a conceptual framework of everyday instructional situations which can assist researchers and practitioners in a more strategic approach of the multidisciplinary literatures related to instruction and learning with technology, as well as providing them with a framework for describing and evaluating any instructional situations. Mehlenbacher *et al.* (2005) propose the following five dimensions of all instructional situations, as shown in Table I.

It seems that there is no published empirical data regarding its effectiveness and efficiency in comparison with other proposed heuristics or techniques.

4. Method

The unit of analysis in this study is an asynchronous e-learning course. Asynchronous e-learning courses are still considered the main building block of the e-learning landscape. A typical commercial asynchronous e-learning course (see acknowledgments) was used and evaluated in this study by two expert evaluators. The course was on "Internet Marketing" and it contained four main learning modules while the total length in e-learning hours was six. The course followed a typical online tutorial model with some theory, examples, drill and practice exercises, self-tests etc.

Dimensions of all instructional situations	Heuristics
Learner background and knowledge	Accessibility
	Customisability and maintainability
	Error support and feedback
	Navigability and user movement
	User control, error tolerance, and flexibility
Learner tasks and activities	*Instructional content*
	Completeness
	Examples and case studies
	Readability and quality of writing
	Relationship with real-world tasks
	Interaction display
	Aesthetic appeal
	Consistency and layout
	Typographic cues and structuring
	Visibility of features and self-description
Social dynamics	Mutual goals and outcomes
	Communication protocols
Instructor activities	Authority and authenticity
	Intimacy and presence
Learning environment and tools	Help and support documentation
	Metaphors and maps
	Organization and information relevance
	Reliability and functionality

Source: Mehlenbacher *et al.* (2005)

Table I.
Dimensions of instructional situations and heuristics

The heuristic evaluation was conducted by two reviewers that have considerable experience in usability evaluation as well as in e-learning; thus they can be characterized as "double experts" according to Nielsen (1992). The goal of the heuristic evaluation was to identify as many usability problems (UPs) as possible, and then to match these with the heuristics of the two heuristic sets. In order to achieve this:

(1) Each reviewer worked alone and afterwards the results were consolidated.

(2) The reviewers first made a free exploration to the e-learning system in order to get familiar with it and took notes in free form; this free exploration lasted for about two hours for each reviewer. This explorative inspection helped the reviewers to get a grip of the basic functions, options and operations available to users as well as to form a general view about general design approach followed for this e-learning system. Furthermore, this stage was useful to identify a number of typical user tasks that should be inspected in more detail.

(3) Then, the reviewers identified and documented usability problems throughout the e-learning system (which took about four-to-five hours) according to a common documentation format. This process included both the inspection of certain important user tasks as well as the identification of more general usability problems.

The structured problem report format (Cockton and Woolrych, 2001) was used for documentation of usability problems. This includes the following descriptors for any usability problem:

(1) A numeric identifier of the problem.
(2) A short description of the problem (and likely difficulties for the user).
(3) Specific context (defined as location of the problem in the interface): we identified this generically in terms of affecting the:
    - structure of the system, including layout, navigation and basic help functions like print, help, etc;
    - e-learning content, including actual content for study, exploration, examples, exercises, and so on; and
    - global, i.e. affecting both structure and e-learning content.
(4) The heuristic(s) used, i.e. any heuristic belonging to the two HS selected.
(5) The severity rate: severe, moderate and minor.

Thus the reviewers followed a common, simple protocol for conducting their evaluation, which is generally required for heuristic evaluations and can make things easier for the integration and interpretation of the results.

5. Results

5.1 Overview

Table II shows an overview of the usability problems found by the heuristic evaluation along with severity ratings and their generic context. Overall, the reviewers felt that a lot of good technical work has been put on to set up this environment (i.e. all major Web technologies were present including HTML, CSS, Javascript and Flash); however several problems were identified regarding usability, accessibility and instructional design issues.

Examples of usability problems found included: too many pop-up windows that disorient users, e-learning content poor in quality that did not cite important scientific papers or documents and not including many examples or case studies and finally the e-learning content was not presented meaningfully according to instructional design guidelines. In addition there are no communication tools available while there is also a lack of authentic resources (such as real data archives, job aids etc). The aforementioned problems are typical of many other e-learning environments that rely heavily on technical development and put aside issues related to pedagogical usability, accessibility and other particular e-learning design issues.

Total number of UPs	76	Per cent
Severity		
Severe	54	71
Moderate	20	26
Minor	2	3
Context		
Severe	30	39
Moderate	20	26
Minor	26	34

Table II.
Total number of usability problems found (UPs), severity and context

5.2 Heuristic sets matched to usability problems: coverage, distribution and redundancy
The matching of heuristic sets to usability problems found is discussed in terms of:

- Coverage, i.e. the degree to which the heuristic sets "include" the usability problems identified. A high coverage of all usability problems identified by a heuristic set for a particular evaluation study indicates that the set is inclusive and does not leave important aspects of usability for e-learning out of its scope.
- Distribution, i.e. the degree to which each heuristic "gathers" a considerable amount of usability problems. Heuristic sets summarise the most important, general principles that are appropriate for the design and evaluation of e-learning systems. If some heuristics tend to gather the large majority of usability problems identified, then it is possible that these should be refined to more specific heuristics, or that the remaining heuristics are not generic enough.
- Redundancy, i.e. the degree to which usability problems appear relevant to more than one heuristic. Heuristic sets summarise distinct principles for the design and evaluation of e-learning systems. If a considerable number of usability problems fit well more that one heuristic, then it is possible that these heuristics are not distinct and allow broader interpretations.

Table III presents an overview of the match of heuristic sets to usability problems indicating the values of the above metrics. With regard to coverage, we have seen that both heuristic sets exhibit high coverage. This is certainly a desirable attribute for any heuristic set (HS). HS no. 1 (Reeves *et al.*, 2002) has incorporated 72 out of 76 (95 per cent) of the usability problems identified. HS no. 2 also exhibits a particularly high coverage of 71 out of 76 (93 per cent) of the usability problems found. Thus, it can be argued that these two sets of heuristics are quite generic and can provide useful guidance to practitioners for design and evaluation of e-learning systems in this respect. The inadequacy of these heuristic sets was quite small (5 per cent and 7 per cent respectively), which is quite encouraging especially if we take into account the relatively large number of usability problems identified.

With regard to distribution, we found some interesting results: for both heuristic sets, there are some heuristics that have "attracted" a large portion of usability problems, while other heuristics attracted too few and some even not a single usability problem! Specifically, for HS no. 1 the most important heuristics in this respect were: "visibility of system status": 22 per cent of usability problems; "interactivity": 16 per cent; and "learning design": 22 per cent. For HS no. 2, the heuristics that have attracted the most usability problems were "accessibility": 25 per cent; navigability: 13 per cent; user control, error tolerance and flexibility: 24 per cent; "readability and quality of writing": 14 per cent. Indeed, most of these heuristics are important usability principles that can be found in all traditional HCI (Human-Computer Interaction) textbooks, while in particular (Web) accessibility is a research field in its own right. Furthermore, for HS no. 2, it can be observed that the following heuristic groups attract the huge portion of usability problems for this evaluation: "learner background and knowledge" (64 per cent), instructional content (34 per cent), interaction display (24 per cent). In contrast to these heuristic groups the rest attract very few usability problems.

The fact that a high distribution is observed to a few heuristics is a strong indication that these heuristics may be quite generic – at least with respect to the rest of the heuristic set. Therefore, there may be a need to further specify these to provide

No.	Reeves *et al.* 2002 (HS no. 1)	UPs	Per cent	No.	Mehlenbacher *et al.* 2005 (HS no. 2)	UPs	Per cent
1	Visibility of system status	17	22	*1*	*Learner background and knowledge*	49	64
2	Match between system and the real world	2	3	1.1	Accessibility	19	25
3	Error recovery and exiting	3	4	1.2	Customizability and maintainability	2	3
4	Consistency and standards	3	4	1.3	Error support and feedback	0	0
5	Error prevention	0	0	1.4	Navigability and user movement	10	13
6	Navigation support	7	9	1.5	User control, error tolerance, and flexibility	18	24
7	Aesthetics	3	4	*2*	*Social dynamics*	5	7
8	Help and documentation	1	1	2.1	Mutual goals and outcomes	4	5
9	Interactivity	12	16	2.2	Communication protocols	1	1
10	Message Design	8	11	*3*	*Instructional content*	26	34
11	Learning Design	17	22	3.1	Completeness	8	11
12	Media Integration	5	7	3.2	Examples and case studies	4	5
13	Instructional Assessment	3	4	3.3	Readability and quality of writing	11	14
14	Resources	5	7	3.4	Relationship with real-world tasks	3	4
15	Feedback	4	5	*4*	*Interaction display*	18	24
UPs that match more than one heuristic (redundancy)		15	20	4.1	Aesthetic appeal	5	7
Ups that do not fit well in any heuristic (N/A)		4	5	4.2	Consistency and layout	1	1
				4.3	Typographic cues and structuring	3	4
				4.4	Visibility of features and self-description	9	12
				5	*Instructor activities*	0	0
				5.1	Authority and authenticity	0	0
				5.2	Intimacy and presence	0	0
				6	*Environment and tools*	5	7
				6.1	Help and support documentation	2	3
				6.2	Metaphors and maps	1	1
				6.3	Organization and information relevance	1	1
				6.4	Reliability and functionality	1	1
				UPs that fit in more than one heuristic		30	39
				Ups that do not fit well in any heuristic		5	7

Table III.
Heuristic sets matched to usability problems (UPs)

more useful guidance to practitioners. On the other hand, there were some heuristics that 'attracted' too few usability problems (3), like "error prevention" and "help and documentation" for HS no. 1 and "error support and feedback" and "communication protocols for HS no. 2. Certainly, it needs to be noted that distribution is affected also by the particular application under analysis and the evaluators" experience, e.g. this course presented several accessibility problems – another course that would be accessible would certainly not have exhibited that many accessibility problems.

With regard to redundancy, we have found that the HS no. 1 (Reeves *et al.*, 2002) exhibited better results (i.e. lesser redundancy) than HS no. 2 (Mehlenbacher *et al.*, 2005). More specifically, a 20 per cent of usability problems were reported for more than one heuristic for HS no. 1, while a 39 per cent of usability problems were reported for more than one heuristic for HS no. 2. This means that for these usability problems it was not particularly straightforward to match them with a single guideline. This is also affected by the particular e-learning application and the evaluators' experience; however there is a notable difference among the two heuristic sets that needs to be further explored.

5.3 Heuristic sets matched to usability problems, including context and severity

Another important dimension of the analysis of heuristic sets is the way by which they capture the context and severity of a usability evaluation. The context was specified in a generic manner and was related to the location of the usability problem in the system; we used a three-scale schema to specify context (global, structure, e-learning content). Severity is related to the degree of importance of any usability problem according to the evaluators; as noted above we have used a three-scale schema to denote severity (severe, moderate, minor). Tables IV and V show the overall picture of heuristic sets when matched to usability problems including context and severity ratings.

With respect to context, it has been observed that HS no. 1 (Reeves *et al.*, 2002) tends to provide heuristics that focus on some aspects of the general context of e-learning applications:

(1) The first eight heuristics (which are the same to Nielsen's ten heuristics of his generic method of heuristic evaluation) tend to be related to usability problems that are concerned with the (more general) structural aspects of the application, such as for example: navigation, clear exit points, error recovery and prevention, and so on. Notably, only three out of 31 usability problems related to the e-learning content have been specified by these first eight heuristics.

(2) The following seven heuristics are clearly related to the e-learning content, notable interaction design issues, learning design, assessment and so on. Thus HS no. 1 presents a clear thematic distinction among its proposed heuristics.

With respect to the specification of context of the HS no. 2 (Mehlenbacher *et al.*, 2005) we can make the following observations:

(1) First, the heuristic group of "learner background and knowledge" attracts a large percentage of usability problems that are related to the structural elements of the e-learning environment. This is also the case with "environment and tools", however the problems identified in this group for this evaluation were quite few. Therefore it seems that learner's background and knowledge is

Table IV.
Heuristic set no. 1 (Reeves *et al.*, 2002) matched to Usability Problems (UPs) including context and severity

No.	Reeves *et al.* 2002 (HS #1)	UPs	%	Context			Severity		
				Global	Structure	E-learning content	Severe	Moderate	Minor
1	Visibility of system status	17	22	8	8	1	15	1	1
2	Match between system and the real world	2	3		1	1	1	1	
3	Error recovery and exiting	3	4		2	1		3	
4	Consistency and standards	3	4	2	1		3		
5	Error prevention	0	0						
6	Navigation support	7	9	3	4		3	4	
7	Aesthetics	3	4	3			1	1	1
8	Help and documentation	1	1	1			1		
9	Interactivity	12	16	7	1	4	11	1	
10	Message Design	8	11	3	2	3	7	1	
11	Learning Design	17	22	7	1	9	14	3	
12	Media Integration	5	7	2		3	4	1	
13	Instructional Assessment	3	4			3	2	1	
14	Resources	5	7	1		4	4	1	
15	Feedback	4	5	1		2	3	1	
	Total	90		38	20	31	69	19	2
				42%	22%	34%	77%	21%	2%

Table V.
Heuristic set no. 2 (Mehlenbacher *et al.*, 2005) matched to usability problems (UPs), context and severity

#	Mehlenbacher *et al.* 2005	UPs	%	Context			Severity		
				Global	Structure	E-learning content	Severe	Moderate	Minor
1	*Learner background and knowledge*	49	64	21%	19%	7%	30%	16%	2%
1.1	Accessibility	19	25	10	6	3	13	5	1
1.2	Customizability and maintainability	2	3		2			1	1
1.3	Error support and feedback	0	0						
1.4	Navigability and user movement	10	13	5	5		7	3	
1.5	User control, error tolerance, and flexibility	18	24	7	7	4	11	7	
2	*Social dynamics*	5	7	3%	0%	2%	4%	1%	0%
2.1	Mutual goals and outcomes	4	5	3		1	4		
2.2	Communication protocols	1	1			1		1	
3	*Instructional content*	26	34	6%	0%	19%	21%	4%	0%
3.1	Completeness	8	11	3		5	6	2	
3.2	Examples and case studies	4	5			4	3	1	
3.3	Readability and quality of writing	11	14	2		9	10	1	
3.4	Relationship with real-world tasks	3	4	1		2	3		
4	*Interaction display*	18	24	6%	7%	5%	12%	5%	1%
4.1	Aesthetic appeal	5	7	2	1	2	3	1	1
4.2	Consistency and layout	1	1	1				1	
4.3	Typographic cues and structuring	3	4		1	2	3		
4.4	Visibility of features and self-description	9	12	3	5	1	6	3	
5	*Instructor activities*	0	0	0%	0%	0%	0%	0%	0%
5.1	Authority and authenticity	0	0						
5.2	Intimacy and presence	0	0						
6	*Environment and tools*	5	7	2%	3%	0%	3%	2%	0%
6.1	Help and support documentation	2	3		2		2		
6.2	Metaphors and maps	1	1	1				1	
6.3	Organization and information relevance	1	1	1				1	
6.4	Reliability and functionality	1	1		1		1		

mainly defined in terms of computer literacy rather than other type of experience or background in e-learning.

(2) Furthermore, HS no. 2 includes a heuristic group of "instructional content" that indeed has largely attracted usability problems related to e-learning content. In general, we can argue that the attribution of specific heuristics to heuristic groups makes sense and can be useful to practitioners for design and evaluation of e-learning systems. On the other hand the remark that some groups have attracted too few usability problems remains and needs to be further investigated with other case studies.

(3) Finally, both heuristic sets exhibited a typical distribution of severity: the large majority of usability problems were characterized as severe and there was not a noticeable heuristic within both HS no. 1 and HS no. 2 that tended to provide minor problems. This is an indication that all heuristics of both heuristic sets can be important for a heuristic evaluation

6. Discussion and future work

Heuristic evaluation is a significant generic method for usability inspection and it has not been extensively explored in e-learning systems evaluation. In this paper we have presented a comparative evaluation of two heuristic sets for e-learning systems proposed by Reeves *et al.* (2002) and Mehlenbacher *et al.* (2005). The method of comparison involved the heuristic evaluation of an asynchronous e-learning course by two reviewers and the matching of usability problems identified (along with severity and context information) to the heuristic sets.

In general, both heuristic sets exhibited high coverage, i.e. they achieved in including the vast majority of usability problems identified. Thus, we feel that both heuristic sets are useful for the heuristic evaluation of e-learning applications.

Both heuristic sets have not exhibited a satisfactory distribution of usability problems. In other words, for both heuristic sets there are some heuristics that attracted a large portion of usability problems, while other heuristics attracted too few (and some not even a single) usability problem(s). Indeed, some heuristics are well acknowledged usability principles, while others are specific guidelines. Therefore, we suggest that there is a need for better orientation and organization of the level of abstraction for both heuristic sets.

With respect to redundancy, we have found that the HS no. 1 (Reeves *et al.*, 2002) exhibited better (i.e. lesser redundancy) than HS no. 2 (Mehlenbacher *et al.*, 2005). In other words, when using the HS no. 2 it was not particularly straightforward to match them with a single guideline, which was not so much of a problem for HS no. 1. Therefore, we suggest that HS no. 2 could be further developed in order to provide more distinct heuristics that can be directly matched to usability problems commonly found in e-learning applications.

When we examined the heuristic sets with respect to the context of usability problems identified, we saw that HS no. 2 was better designed in terms of specifying the context for each of its heuristics which is apparently quite helpful for usability evaluators. HS no. 1 does not provide contextual information of this sort, despite that it seems to consist of two generic types of guidelines: those concerned with usability in general (web usability) and those related to e-learning systems in particular

(pedagogical usability). Therefore, we suggest that HS no. 1 could be further developed to better describe the specific context of e-learning systems.

The interpretation of the results of any comparative heuristic evaluation mainly depends on the heuristic sets used, the method employed for the conduction of the evaluation, the particular features of the system and the evaluators' experience.

- With respect to the method used, we followed the guidelines of Nielsen(1992) for the conduction of usability evaluations, keeping the number of evaluators to only two in order to better process the comparison of results. However, we acknowledge that a larger number of evaluators would contribute to greater numbers of usability problems found and possibly to more refined recommendations.
- With regard to the system under analysis, there is a need to perform a considerable number of studies including a larger number of e-learning systems in order to strengthen the aforementioned results and produce more refined recommendations.
- With regard to the issue of evaluators' experience, of particular interest for comparative evaluations is the work of Jacobsen *et al.* (1998) who investigate the evaluator effect in usability evaluations; i.e. the fact that multiple evaluators evaluating the same interface with the same user evaluation method detect markedly different sets of problems. We followed a consistent method for the conduction and documentation of the usability evaluation in order to minimize this effect.

The work presented in this paper shows the application of two heuristic sets in a usability evaluation of an e-learning application. Furthermore, it provides a comparison of the two heuristic sets according to a set of criteria and provides a first set of suggestions regarding further development of these heuristic sets. As a future prospect we plan to further investigate e-learning evaluation by comparing the results of the two heuristic sets to user testing methods. In addition future work will employ comparative evaluation studies that will include a larger number of evaluators and more criteria for evaluation such as realness, validity, thoroughness and effectiveness (Koutsabasis *et al.*, 2007). Another future work direction is to augment the existing heuristic protocols so as to cover more contemporary types of e-learning applications, such as epistemic games and e-learning 2.0 applications.

References

Albion, P.R. (1999), "Heuristic evaluation of educational multimedia: from theory to practice", *Proceedings of 16th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, ASCILITE*.

Ardito, C., Costabile, F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, Z.T. and Rossano, V. (2006), "An approach to usability evaluation of e-learning applications", *Univ Access Inf Soc.*, Vol. 4, pp. 270-83.

Ardito, C., De Marsico, M., Lanzilotti, R., Levialdi, S., Rossano, V. and Tersigni, M. (2004), "Usability of e-learning tools", *Proceedings of the Working Conference on Advanced Visual Interfaces AVI '04, Gallipoli*, pp. 80-4.

Cockton, G. and Woolrych, A. (2001), "Understanding inspection methods: lessons from an assessment of heuristic evaluation", in Blandford, A. and Vanderdonckt, J. (Eds), *People & Computers XV*, Springer, Berlin.

Cordes, R.E. (2001), "Task-selection bias: a case for user-defined tasks", *International Journal of Human-Computer Interaction*, Vol. 13 No. 4, pp. 411-9.

Dringus, L. (1995), "Interface issues associated with using the internet as a link to online courses", *Journal of Interactive Instruction Development*, pp. 16-20.

Dringus, L.P. and Cohen, M.S. (2005), "An adaptable usability heuristic checklist for online courses", *Proceedings of Frontiers in Education, 35th Annual Conference*.

Dumas, J.S. and Redish, J.C. (1993), *A Practical Guide to Usability Testing*, Ablex Publishing, Norwood, NJ.

Georgiakakis, P., Papasalouros, A., Retalis, S., Siassiakos, K. and Papaspyrou, N. (2005), "Evaluating the usability of web-based learning management systems", *THEMES in Education*, Vol. 6 No. 1, pp. 45-59.

Hartwig, R., Schön, I. and Herczeg, M. (2003), "Usability engineering in computer aided learning contexts results from usability tests and questionnaires", in Jacko, J. and Stephanidis, C. (Eds), *HCI Human Computer Interaction – Theory and Practice (Part I)*, Lawrence Erlbaum Associates Publishers, London, pp. 946-50.

Hornbaek, K. (2006), "Current practice in measuring usability: challenges to usability studies and research", *International Journal of Human-Computer Studies*, Vol. 64, pp. 79-102.

Hornbæk, K. and Frøkjær, E. (2008), "Comparison of techniques for matching of usability problem descriptions", *Interacting with Computers*, Vol. 20 No. 6, December, pp. 505-14.

Hvannberg, E.T., Law, E.L. and Larusdottir, M.K. (2006), "Heuristic evaluation: comparing ways of finding and reporting usability problems", *Interacting with Computers*, doi: 10.1016/j.intcom.2006.10.001

Jacobsen, N.E., Hertzum, M. and John, B.E. (1998), "The evaluator effect in usability tests", *in Summary Proceedings of the ACM CHI 98 Conference, ACM Press, New York, NY*, pp. 255-6.

Jeffries, R., Miller, J.R., Wharton, C. and Uyeda, K.M. (1991), "User interface evaluation in the real world: a comparison of four techniques", *Proceedings of CHI'91: Human Factors in Computing Systems, ACM Press/Addison Wesley, New York, NY*, pp. 119-24.

Koutsabasis, P., Spyrou, T. and Darzentas, J. (2007), "Evaluating usability evaluation methods: criteria, method and a case study", *12th International Conference on Human-Computer Interaction, Beijing, China, 2007*, Lecture Notes in Computer Science, No. 4550, Springer, New York, NY.

Ling, C. and Salvendy, G. (2005), "Extension of heuristic evaluation method: a review and reappraisal", *Ergonomia IJE&HF*, Vol. 27 No. 3, pp. 179-97.

Mehlenbacher, B., Bennett, L., Bird, T., Ivey, M., Lucas, J., Morton, J. and Whitman, L. (2005), "Usable e-learning: a conceptual model for evaluation and design", *Proceedings of HCI International 2005: 11th International Conference on Human-Computer Interaction, Volume 4 – Theories, Models, and Processes in HCI, Mira Digital, Las Vegas, NV*, pp. 1-10.

Nguyen, T. and Chang, V. (2006), "A user-centred personalised e-learning system", in Shackleton, P. (Ed.), *We-B Conference 2006: e-Business: How Far Have We Come?, Victoria University of Technology, Melbourne, Australia*, pp. 192-9.

Nielsen, J. (1992), "Finding usability problems through heuristic evaluation", *Proceedings of CHI Conference on Human Factors in Computing Systems, ACM, New York, NY*, pp. 373-80.

Nielsen, J. (1994), "Enhancing the explanatory power of usability heuristics", *Proceedings of ACM CHI'94 Conference, Boston, MA, April 24-28*, pp. 152-8.

Nielsen, J. and Molich, R. (1990), "Heuristic evaluation of user interfaces", *Proceedings of ACM CHI'90 Conference, ACM, Seattle, WA*, pp. 249-56.

Nokelainen, P. (2006), "An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students", *Educational Technology & Society*, Vol. 9 No. 2, pp. 178-97.

Phillips, D.C. (1995), "The good, the bad, the ugly: the many faces of constructivism", *Educational Researcher*, Vol. 24 No. 7, pp. 5-12.

Pipan, M., Arh, T. and Jerman-Blažič, B. (2006), "The evaluation cycle management – model for the assessment of the usability and applicability of learning management systems. Is information technology shaping the future of higher education?", *Proceedings of the 12th International Conference of European University Information Systems, EUNIS 2006, Estonia*, pp. 325-32.

Quinn, C.N. (1996), "Pragmatic evaluation: lessons from usability", *Proceedings of 13th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*.

Reeves, T., Benson, L., Elliott, D., Grant, M., Holschuh, D., Kim, B., Kim, H., Lauber, E. and Loh, S. (2002), "Usability and instructional design heuristics for e-learning evaluation", *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications AACE, Charlottesville, VA*, pp. 1615-21.

Silius, K. and Tervakari, A-M. (2003), "An evaluation of the usefulness of web-based learning environments. The evaluation tool into the portal of Finnish virtual university", in Peñarrocha, V. (Ed.), *mENU 2003 – International Conference on University Networks and E-learning, Valencia, 8-9 May*.

Soloway, E., Guzdial, M. and Hay, K.E. (1994), "Learner-centered design: the challenge for HCI in the 21st century", *Interactions*, April, pp. 36-48.

Squires, D. and Preece, J. (1999), "Predicting quality in educational software: evaluating for learning, usability and the synergy between them", *Interacting with Computers*, Vol. 11 No. 5, pp. 467-83.

Zaharias, P. and Poulymenakou, A. (2009), "Developing a usability evaluation method for e-learning applications: beyond functional usability", *International Journal of Human-Computer Interaction*, Vol. 25 1, January, pp. 75-98.

Zaharias, P. (2007), "Heuristic evaluation in e-learning context: selecting the appropriate tasks and reporting usability problems", *Proceedings of International Conference on E-Learning – ICEL07*.

Further reading

Ssemugabi, S. and De Villiers, R. (2007), "A comparative study of two usability evaluation methods using a web-based e-learning application", *Proceedings of the 2007 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries. Port Elizabeth, South Africa*, pp. 132-42.

About the authors

Panagiotis Zaharias is currently a Visiting Assistant Professor at the Department of Computer Science at the University of Cyprus. He holds a PhD degree in Information Systems (specialization on human computer interaction, thesis title: "A usability evaluation method of e-learning courses") from the Department of Management Science & Technology of Athens University of Economics and Business (2004). His main research interests are focused on usability and user experience evaluation methods, e-learning design, virtual worlds in education and game-based learning. Panagiotis Zaharias is the corresponding author and can be contacted at: pz@aueb.gr

Panayiotis Koutsabasis has a PhD in Information Systems. He is a Lecturer at the Department of Product and Systems Design Engineering, University of the Aegean. His main area of research is human-computer interaction with particular emphasis in human-centred design methods, adaptive user interfaces, design for all and natural interaction in applications such as e-commerce, e-learning and design support.