Ergonomia IJE&HF, 2005, Vol. 27, No. 3, 179–197

Extension of heuristic evaluation method: a review and reappraisal

Chen Ling[1], **Gavriel Salvendy**[1,2]

[1] School of Industrial Engineering, Purdue University, West Lafayette, Indiana, USA
[2] Department of Industrial Engineering, Tsinghua University, Beijing, PR China

Abstract

As one of the major discount methods, the heuristic evaluation method (HE) is the most commonly used usability inspection method. We introduced the history and procedure of this method, as well as its strengths and weaknesses. We then reviewed the applications of this method to different Human-Computer Interaction systems with the adapted heuristic sets. We also reviewed many studies that extended the traditional HE method in different ways. Finally, the paper ends with a reappraisal of these extension methods and future research direction to improve the HE method.

Key words: heuristic evaluation, usability inspection, extension method, adapted heuristics

Address for correspondence:
Gavriel Salvendy, Grissom Hall, Room 263, 315 N. Grant Street,
School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA
e-mail: salvendy@ecn.purdue.edu

There are many methods for the evaluation of information technology products and services (Nielsen, Mack 1994; Lewis 2006) for ease, joyful and productive use. According to a major survey result, heuristic evaluation (HE) method is currently the most used usability method (Rosenbaum et al. 2000). It is a widely accepted and applied in both academia and industry. To achieve better evaluation result with the HE methods on various information technology systems, many studies have been conducted to extend the traditional HE method in different ways. A comprehensive review of studies on extension to HE method is provided.

1. History and procedure

The concept of "discount usability method" (Nielsen 1993) has been around for more than a decade. The asset of discount usability methods is to save time and cost of carrying out a usability study while achieving satisfactory study results. Heuristic evaluation method is one of the main discount methods because it can effectively detect usability problems with limited time and resources (Nielsen 1994b). Nielsen and Molich invented the HE method in 1990 (Molich, Nielsen 1990; Nielsen, Molich 1990). It is a usability inspection method whereby a set of evaluators produces lists of usability problems in a user interface by going through it and noting deviations from accepted usability principles (Nielsen, Phillips1993). These accepted usability principles are also called heuristics. The evaluation of interface used to be difficult and time-consuming due to the intimidating number of guidelines (in thousands) to observe. Nielsen and Molich (1990) cut down the complexity of the extensive collections of guidelines by two orders of magnitudes, and derived ten "golden rules" or heuristics. These heuristics were chosen based on Nielsen and Molich's understanding of typical problem areas of usability, as well as an informal consideration of existing guidelines. They described their method as "the most general of the usability inspection methods and is also the easiest to learn and apply" (Nielsen, Molich1990). The original list of usability heuristics are listed here.

- Simple and natural dialogue
- Speak the user's language
- Minimize the user's memory load
- Consistency
- Feedback
- Clearly marked exits
- Shortcuts
- Precise and constructive error messages
- Prevent errors
- Help and documentation

Nielsen (1994a) later performed a more formal study on heuristics. He chose 101 usability principles, which includes the original set of heuristics listed above, as well as six other collections of published principles or guidelines. He studies how well these principles can account for the 249 usability problems found during the evaluation of 11 interactive systems. He attempts to pick the principles that provided the best explanation of the usability problems. The result of a principal component analysis indicates that seven factors could account for 30% of the variability of the usability problems. The seven factors formed the basis for a revised set of 10 heuristics with maximum explanatory power. Another three heuristics were added by Nielsen based on his own experience. The commonly used ten heuristics for interactive systems are listed as follows (Nielsen 1994b).

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation

During the heuristic evaluation, the evaluators decide on their own how they want to proceed with the evaluation. It is generally recommended that they go through the interface at least twice. The first pass is intended to get a feel for the flow of the interaction and the general scope of the system. The second pass then allows the evaluator to focus on specific interface elements. Heuristic evaluators use their judgment to determine whether an interface violates any of the heuristics (Nielsen, 1994b).

Different people find different usability problems and give different severity ratings to usability problems using HE (Nielsen 1994b; Hertzum, Jacobsen 2001). Therefore, it is necessary to involve multiple evaluators in HE. It is recommended to use around five evaluators because it can cover around 75% of total usability problems and have the highest benefit-to-cost ratio (Nielsen, Landauer 1993; Nielsen 1994b).

HE method can find both major and minor usability problems, but most problems found are minor ones (Nielsen, 1994b). It has higher thoroughness than cognitive walkthrough method because it finds more intermediate and minor problems (Sears 1997). But it can also find more false-positive problems that will not actually occur with real users and need to be eliminated (Sears 1997). Studies show that the use of HE early in the design process tends to miss certain classes of problems, such as those arising from perceptual-motor slips (Mack, Montaniz 1994), or missing functionality (Nielsen 1992). This might be compensated by directing the inspectors to pay particular attention to these areas. When compared to user testing, HE tends to miss task-based problems, whereas the user testing tends to miss interface feature related problems because heuristic evaluators were not absorbed in using the system to perform a task as the users in the user testing (Doubleday et al.1997). HE usually identifies the cause of the problem while end user testing may indicate the symptom of the problem (Doubleday et al. 1997). Another empirical study shows that HE is more effective in identifying usability problems associated with skill-based and rule-based levels of performance whereas user testing is more effective in identifying usability problems associated with knowledge-based level of performance (Fu et al. 2002). Since HE and usability testing complement each other by identifying different sets of usability problems, it is suggested that they should be used together to find more comprehensive usability problem set. The common suggestion is to apply HE first to get the "lower hanging fruit", and then perform usability testing to clarify the rest of the problem areas (Nielsen 1994b).

2. Previous studies

HE method has been widely used to evaluate and improve the usability of many systems. When first developed, HE was applied to the many low fidelity prototypes of telephone system such as the screen dumps, written specification of information system, and voice response system (Nielsen, Molich 1990). Nielsen then applied the method to other telephone system such as telephone operated interface in banking system (Nielsen 1992), and the integrating system for internal telephone company use (Nielsen 1994b). After its effectiveness in detecting usability problems has been established by research (Jeffries et al. 1991; Nielsen 1992), it was widely applied to many different computing systems, such as the hypermedia browser (Connell, Hammond 1999), digital libraries (Blandford et al. 2004), World Web Prototype (Levi, Conrad 1996), and museum web site (Bendoly, Goldman 2003) etc. It continues to be used for newly emerged information systems and devices, such as Palm Pilot personal organizer (Slavkovic, Cross 1999), laboratory teaching tool (Avouris et al. 2001), and brokerage platform for E-learning (Law, Hvannberg 2002). It was also used to analyze the learning process of the object-oriented languages (Warren, 2004) and the programming practice in integrated development environment of C++ (Kline 2002).

HE method has been used beyond the traditional information technology industries. It was applied to improve the musical related hardware such as electrical guitar amplifier (Fernandes, Holmes 2002). It helps to find usability problem with an acoustic fishing display (Mills 1995). It improved the usability of many industrial hypermedia applications used by shop floor operators (Fakun, Greenough 2002). Health care industry start to benefit from using it by correcting usability problems with the process of Internet Telemedicine (Lathan et al. 1999), the medical devices such as infusion pumps (Zhang et al. 2003), and healthcare information systems (McGrow et al. 2004).

To better address usability issues in computing systems from different domains, specific heuristic sets can be developed. These adapted heuristics sets will be discussed in detail in section 4.1.

3. Strengths and weaknesses

Heuristics evaluation method has both strengths and weaknesses. As a discount method, HE method has the advantage of effectively identify the usability problems, and identify the most serious problems (Jeffries et al. 1991). At the same time, it is easy to learn and fast and cheap to apply. Because it can be applied to low fidelity prototypes, it become an important method early in the development (Nielsen 1994b). Its evaluation procedure is relatively informal due to its free-form structure. Only a small number of evaluators are needed. Heuristics has key benefits of concise, memorable, meaningful and insightful (Paddison, Englefield 2003). The HE method can be customized to different domains by developing domain-specific heuristics. (Nielsen 1994b). All these strength made HE the most used usability method (Rosenbaum et al. 2000).

Despite its popularity, HE method has many drawbacks. Though HE is said to be a systematic inspection of the user interface design for usability (Mack, Nielsen 1994; Nielsen, Mack 1994), its evaluation procedure is loosely structured (Jeffries et al. 1991). Studies show "substantial unexplained variability in performance from one evaluation to the next" (Nielsen1994b). The past literature listed at least four weaknesses of the HE method:

(1) HE relies heavily on the expertise of the evaluator both in usability and in domain knowledge (Doubleday et al. 1997). Heuristic valuators can be novice, usability expert and double expert in both usability and domain knowledge (Nielsen 1994b). Double experts can find most problems while novice finds the least. But the experts are always harder to find.

(2) HE lacks discovery resources to discover usability problems. Nielsen noted that the heuristic set is good at explaining existing problems, but he does not know how effective they will be at finding usability problems. Heuristics are not always able to deeply guide evaluators (Doubleday et al.1997). Cockton and Woolrych (2001) developed a Discovery and Analysis Resources (DR-AR) model to differentiate two distinct evaluator's activities: discovery and problem analysis. Effectiveness of a method depends on the resources that the evaluator have for these two activities. HE method lacks both discovery and analysis resources. The HE method does little to prepare evaluators for inspection. The essential discovery resource that HE lacks is a focus on task execution and complex domain goals that are the origin of subtle interaction problems (Cockton, Lavery, Woolrych 2003).

(3) HE method produces a large number of false positive usability problems which require evaluators to spend extra time to eliminate (Miller, Jeffries 1992). With only a list of heuristics providing guidance, the HE process is unstructured and does not focus on user tasks (Sears 1997). Because HE

does not have much control over its process, the less experienced evaluators may lose focus on the users. They may focus on issues that do not impact users, and therefore may find unreal usability problems (Sears 1997). The HE method lacks the analysis resource because it lacks a respect for user's intelligence and an understanding of display-based interaction that eliminates logically possible problems as empirically improbable (Cockton, Lavery, Woolrych 2003).

(4) HE method doesn't provide a systematic way to generate fixes to usability problems, or a way to assess the probable quality of any redesign, thus, it identifies usability problems without providing direct suggestions for how to solve them (Nielsen 1994b)

4. Extension of heuristic evaluation method

To produce a more complete set of usability problems in a system, practitioners have modified HE method in different ways to suit their evaluation needs for different computing systems and try to compensate for the drawbacks of the HE method. The modification usually occurs in three ways, extending the heuristics set, extending the HE method by modifying the evaluation procedure, and extending the HE method with a conformance rating scale.

4.1 Extended heuristic set

HE method was developed and applied mainly for single user, productivity-oriented desktop programs, which were the major computer application in the early 90s. However, with computer technologies getting more integrated into everyday life and new types of human computer interaction emerging, Nielsen's 10 heuristics may not be able to cover usability issues in the new computing systems. For example, mobile systems need to address issues of changing context of use (Vetere et al. 2003). Notification systems and ambient display system need to study the overall performance of the system including both the traditionally studied primary tasks, and the secondary tasks as well (Berry 2003). Teamwork issues involving multiple users rather than single user task work issues need to be addressed by groupware systems (Gutwin 2000; Drury 2001). The goal of the traditional system is usually fast and easy, but for the game system, the goal is easy to learn, but hard to master (Desurvire et al. 2004). Thus, games designs need to intentionally contravene Nielsen's heuristics of prevent errors, but provide possibility of errors for the users instead (Johnson, Wiles 2003). Therefore, Nielsen's ten heuristics are not readily applicable to many new domains with different goals and usability issues. As domain-specific heuristics can be developed to supplement the existing heuristics (Molich, Nielsen 1990; Nielsen 1993; Nielsen, Mack 1994), researchers have derived many adapted heuristic sets to address the typical requirements and problems in different kinds of application domain. Table 1 lists studies that developed new sets of heuristics, the domain that the heuristic set was developed for, a brief description of the heuristics, and the way that they are developed.

These studies cover many types of heuristics sets. Some are intended for use across different technologies (Kamper 2002), and most are adapted domain-specific heuristics intended for specific type of computing systems with different goals and interaction nature. Important design issues that are not covered by the Nielson's traditional heuristics set are emphasized in the heuristics according to the different purposes. For example, to design better playability of games (Desurvire et al. 2004), four aspects need to be addressed: game play, game story, mechanics and usability. Designing a better electronic newspaper needs to stress three aspects, including graphics, general layout, navigation (Mariage, Vanderdonckt 2000).

From these studies, we can see that it generally takes two steps to derive a new set of heuristics: heuristic development and

Table 1. Adapted heuristic sets.

Reference	Application domain	Number and content of heuristics	Developed based on
Kamper 2002	Many different domains	Lead (6) Follow(6) Get out of way (6)	Previous research
Rau, Liang 2003	Websites	Information Design (10) Consistency (2) Navigation (2) Operation (6) Errors (4)	Nielsen's heuristic set and Garzotto et al.1995
Sutcliffe 2001	Website attractiveness	Attractiveness (7) Content (5)	Nielsen's heuristic set and previous research
Paddison, Englefield 2003	Accessibility	9: Provide meaningful and relevant alternatives to non-text elements; Support consistent and correctly tagged navigation; Allow complete and efficient keyboard usage; Ensure appropriate use of standard and proprietary controls; Do not rely on colour alone to code and distinguish; Allow users to control of potential distractions; Allow users to understand and control time restraints; Make certain the Web site is content compatible with assistive technologies.	Previous research
Muller et al. 1998	Participatory	15: Revised Nielsen's set plus Respect the user and his/her skills; Pleasurable experience with they system; Support quality work.	Traditional heuristic set (Nielsen, Molich 1990)
Evans, Sabry 2003	Web-based learning system	9: Making navigation easy; Engaging learner frequently; Allow for reflection; Using a variety of interactions; Using multimedia; Applying what has been taught; Being relevant; Being timely; Grabbing attention.	Previous research
Reeves 2002	E-learning system	15: Revised Nielsen (1994b)'s set plus interactivity; Message design; Learning design; Media integration; Instructional assessment; Resources; Feedback.	Nielsen's heuristic set
Squires, Preece 1999	Educational software	8: Match between designer and learner models; Navigational fidelity; Appropriate levels of learner control; Prevention of peripheral cognitive errors; Understandable and meaningful symbolic representation; Support personally significant approaches to learning; Cognitive error recognition, diagnosis and recovery cycle; Match with the curriculum is evident.	Nielsen's heuristic set and previous research
Greenberg et al. 1999	Shared workspace groupware system	5: Provide centers; Provide awareness; Allow Individual views; Allow people to manage and stay aware of their evolving interactions; Provide a way to organize and relate locales to one another.	Previous research

Table 1 cont. Adapted heuristic sets.

Baker et al. 2002	Shared workspace groupware system	8: Provide the means for intentional and appropriate verbal communication; Gestural communication; Provide consequential communication of an individual's embodiment; Shared artifacts; Provide protection; Management of tightly and loosely-coupled collaboration; Allow people to coordinate their actions; Facilitate finding collaborators and establishing contact.	Previous research
Drury 2001	Synchronous collaborative systems	8: Show identities of people in the workspace; Show the activities of the other participants; Show the locations where the other participants are working; Show the changes made by other participants; Show the goals of other participants; Show the interdependencies of participants' work; Show the extent to which the system supports social rule; Provide clues for users to predict the system's probable future status.	Previous research
Mankoff et al. 2003	Ambient displays	12: Sufficient information design; Consistent and intuitive mapping; Match between system and real world; Visibility of state; Aesthetic and pleasing design; Useful and relevant information; Visibility of system status; User control and freedom; Easy transition to more in-depth information; "Peripherality" of display; Error prevention; Flexibility and efficiency of use.	Nielsen's heuristic set
Berry 2003	Notification system	8: Timely; Reliable; Consistent; Information understandable; Shortcut; Indicate status; Provide context; Allow adjustment.	Evaluation results
Somervell et al. 2003	Large screen information exhibits	8: Appropriate color schemes can be used for supporting information understanding; Layout should reflect the information according to its intended use; Judicious use of animation is necessary for effective design; Use text banners only when necessary; Show the presence of information, but not the detail; Using cyclic displays can be useful, but care must be taken in implementation; Avoid the use of audio; Eliminate or hide configurability control.	Evaluation results
Köykkä et al. 1999	3D multi-user interface	12: modified Nielsen's set plus Real world metaphors have to be clearly understandable; Provide support for orientation, navigation and movement; Avoidance of delays and waiting periods in the performance.	Nielsen's heuristic set and evaluation results
Vetere et al. 2003	Mobile use	Provide locales(6) Provide awareness (5) Provide individual views (3) Interaction trajectories (4) Civic structures (5)	Previous research

Table 1 cont. Adapted heuristic sets.

Desurvire et al. 2004	Playability of games	Game play (16) Game story (8) Mechanics (7) Usability (12)	Previous research
Zhang et al. 2003	Healthcare medical devices	14: Consistency and standard; Visibility of system state; Match between system and world; Minimalist; Minimize memory load; Informative feedback; Flexibility and efficiency; Good error messages; Prevent errors; Clear closure; Reversible actions; Use user's language; Users in control; Help and documentation.	Nielsen's heuristic set and Shneiderman (1998)'s guideline
McGrow et al. 2004	Healthcare information system	24: For example: Timely feedback from the computer; User in control; Easy to move and navigate; Undo unwanted actions at anytime; Escape or exit from the program at any time, etc.	Nielsen's heuristic set
Fakun, Greenough 2002	Industrial hypermedia applications	7: Easy retrace mechanism; Overview, zoom, filter, details search strategy; Information chunks of single concepts; A sense of entire domain and a preview device; All elements have same conventions; Judicious use of colors; Known metaphor.	Previous research
Mariage, Vanderdonckt 2000	Electronic newspaper	Graphics (7) General layout (7) Navigation (4)	Previous research

heuristic validation. In the heuristic development step, researchers come up with a new set of heuristics for use in the HE method. If the new set is developed on the basis of the Nielsen's heuristics (Mankoff et al. 2003, Muller et al.1998), then the common steps include first taking off the not-applicable heuristics from the Nielson's 10 heuristics based on whether the heuristics met the primary goal of the system, then modifying the applicable heuristics to suit the use within the domain, and lastly adding additional heuristics to form the new set. Many studies come up with new set of heuristics without referencing the Nielson's original heuristics. These heuristic sets don't bear much similarity to the original set in structure and content. Regardless of whether the traditional heuristic set was referenced or not, to come up with new heuristics, researchers always need to consult the past literatures, or past evaluation results, and the expert opinions. Two common approaches of deriving new heuristics include the research-based method and the evaluation-based method (Paddison, Englefield 2003). In the research-based method, key points of the domain are identified based on past literatures. For example, Drury (2001) constructed eight heuristics for the synchronous collaborative system based on the six related theories and metaphors for the type of system. In the evaluation-based method, common usability problems with the system are categorized into heuristics. For example, Berry (2003) categorized the common usability problems with the notification system into eight major heuristics to form the heuristic set. Sometimes, second-level heuristics with finer granularity were developed to help evaluators better understand the intent and focus of the heuristics. The whole sets are usually reviewed and modified by domain experts before they are used in further validation experiments.

In the validation phase of the heuristic adaptation, the authors of the newly developed heuristics set will compare its effectiveness with Nielson's original sets by conducting empirical studies (Baker et al. 2002; Berry 2003; Mankoff et al. 2003), or benchmark with user testing results (Desurvire et al. 2004). The adapted set of heuristics can usually obtain better or comparable effectiveness as the Nielsen's original sets because their heuristics fit the evaluated domain better (Baker et al. 2002; Berry 2003; Mankoff et al. 2003).

The adapted heuristics can undergo several rounds of modifications before a final optimal set of heuristics emerge. For example, three out of nine heuristics for web-based learning systems were changed after the evaluation study and discussions (Evans, Sabry 2003). To develop a heuristic set for groupware system, Greenberg et al. (1999) first derived five heuristics based on the Locales framework of social interaction. Later on, the mechanics of collaboration framework was used as the theory basis to address a narrower focus: the shared visual workspace. Then, a new set of eight heuristics was developed (Baker et al. 2001) and empirically studied (Baker et al. 2002) for the groupware system.

After the new heuristics set is developed, the process of applying the HE method usually remains the same to achieve its benefit of "discount methods". However, there can be some additional changes. For example, Vetere et al. (2003) used the heuristic walkthrough procedure with their mobile heuristic set rather than the HE evaluation procedure. In participatory heuristic evaluation (PHE) (Muller et al.1998), three "participatory" heuristics were added and validated to consider the usage context of the system. Rather than just using expert inspectors as in traditional HEs, the PHE stipulates involving users as inspectors for the participatory nature.

4.2 Extension methods with modified evaluation procedure

In an attempt to overcome HE's drawbacks and produce better results, many studies extended the original HE method in several different ways. Table 2 lists seven extended methods with a brief description of the major characteristics and the system the method has been applied to. As these methods extend the HE method in their unique ways, we will describe them in detail in the following.

To give more structure to the traditional HE methods, Sears (1997) created a technique called **heuristic walkthrough** that combines benefits from the HE, cognitive walkthrough and usability walkthrough. While keeping a free-form evaluation run from HE, it adds the task focus of cognitive walkthrough to bring in more structure. The evaluators evaluate the interface in two passes, a task-oriented evaluation guided by four thought-focusing questions derived from cognitive walkthrough, followed by a free-form evaluation guided by heuristics. When compared to HE, the heuristic walkthrough methods produced fewer false positives usability problems. Thus its results have higher validity than HE.

The **HE-plus method** added a contextualized layer called "usability problem profile" to aid the evaluation process (Chattratichart, Brodie 2002). The usability problem profile contains the common usability problem areas associated with the same type of application/product. For example, profile for e-commerce website includes six problem areas: content, graphics, navigation, layout, terminology and matches with user's tasks. In the HE-plus method, a list of the problem areas that constitute "usability problem profile" was given to the evaluators in addition to the list of heuristics. Evaluators are asked to look for problems in the areas given in the usability problem profile while they examine the web site against the heuristic set. It is proven to give results with higher reliability, thoroughness, validity and effectiveness than the HE method when applied to two online shopping websites (Chattratichart, Brodie 2002, 2004). It is reasoned that this is because that the usability problem profile

Table 2. Extension methods to heuristic evaluation.

Reference	Extended methods	Intention	Characteristics	System applied to
Sears 1997	Heuristic walkthrough	To bring in more structure to HE	Task-based evaluation with thought-provoking questions followed by free-form evaluation with heuristics	Educational application
Chattratichart 2002, 2004	HE-plus	To increase evaluation reliability	Usability problem profile	Online shopping website
Garzotto 1998; Matera et al. 2002; Angeli et al. 2003	Systematic usability evaluation (SUE)	To guide novice evaluators; and transfer evaluator's know-how	Abstract tasks; Inspection followed by user testing	Hypermedia CD-ROMs
Kurosu 1997, 1998, 1999	Structured heuristic ealuation method (sHEM)	To avoid overwhelming number of heuristics and help evaluators focus on a limited issues at a same time	Five sub-sessions: ease of cognition, ease of operation and pleasantness, novice/expert, and disabled users.	Portable mini-disk
Desurvire, Thomas 1993	Programmed amplification of valuable experts (PAVE)	To encourage a broader scope of thinking; to improve novice evaluator's performance	Ten perspectives: self, a human factor expert, a cognitive psychologist, a behaviorist, a social/community psychologist, an anthropologist, a Freudian, a health advocate, a worried mother, and a spoiled child.	Flow chart of voice interface
Zhang et al. 1999	Perspective based heuristic evaluation	To give more guidance to evaluation and encourage broader thinking	Three perspectives: novice use, expert use, and error handling; Evaluation procedure for each perspective	Web-based interface
Hornbæk, Frøøkjær 2004	Metaphors of thinking (MOT)	To focus inspection on user's mental activity and to make inspection easily applicable to different devices and use contexts	Five metaphor of human thinking: habit formation is like a landscape eroded by water; thinking as a stream of thought; awareness as a jumping octopus in a pile of rags; utterances as splashes over water; knowing as a building site in progress.	Portal for students at university website

helps the evaluators focus their evaluations on important problem areas.

The **systematic usability evaluation** (SUE) method is especially used for evaluating hypermedia usability (Garzotto et al. 1998; Matera et al.2002; Angeli et al. 2003). It supplies the evaluators with a structured flow of activities by the means of "abstract tasks". An evaluation pattern called abstract tasks describes in detail the activities the evaluators should perform. SUE is believed to facilitate sharing and transferring evaluation know-how among evaluators. Based on the Hypermedia design model (HDM), SUE focuses not only on the surface elements of the system, but also on specific hypermedia aspects such as navigation and information structures, synchronization etc. Another characteristic of the SUE method is conducting user testing after the problems from inspection are fixed to validate subjective aspects such as learning behavior and user satisfaction. Empirical comparison between the SUE method and HE shows that, the SUE method is better than the HE in terms of effectiveness, efficiency and satisfaction (Angeli et al. 2003).

In order that the evaluators are not overwhelmed by the large number of heuristics, Kurosu et al. (1997) developed **structured heuristic evaluation method** (sHEM) by introducing structure into the set of usability heuristics. The usability heuristics (Nielsen, Molich 1990) are split into three sub-categories: ease of cognition, ease of operation and pleasantness. Two additional categories concerning the users includes novice vs. expert, and users with special care. The evaluation is divided into sub-sessions, with each sub-session devoted to finding usability problems concerning heuristics belonging to one of the sub-categories. Each sub-session takes 30 minutes and there is 15 minutes break between sub-sessions. Because within each sub-session, subjects concentrate on finding usability problems with limited amount of heuristics within the range of human memory chunk size (7 plus/minus 2), and several sub-sessions can cover more usability issues than HE, it is believed that this type of evaluation will be more productive than the HE method. The results show that sHEM can find twice as much usability problems than the traditional HE method (Kurosu et al. 1998, 1999).

Programmed amplification of valuable experts (PAVE) method (Desurvire, Thomas 1993) was developed to encourage a broader scope of thinking by including ten perspectives to the evaluation: self, a human factor expert, a cognitive psychologist, a behaviorist, a social/community psychologist, an anthropologist, a Freudian, a health advocate, a worried mother, and a spoiled child. It intends to help novice evaluator reveal as many usability problems as experts by augmenting the evaluator's existing knowledge and stimulating them to think about usability more broadly with the help of different perspectives. In this method, evaluators will study the interface for ten times, each time with one of the ten perspectives. When applied to a flow chart of a voice interface (Desurvire, Thomas 1993), it improved the performance of a novice evaluator by giving less false positives, finding real problems and offering more suggestion for improvements. But this method doesn't improve the performance of the expert evaluators.

Similar to the PAVE approach, Zhang et al. (1999) proposed the **perspective-based method.** The method uses perspectives to focus the evaluator's attention on a specific subset of usability issues during each evaluation session. It asks the evaluators to inspect the interface focusing on one of the three defined perspectives: novice use, expert use, and error handling. Each perspective is associated with a set of inspection questions based on modified "Seven Stages of Action" Model (Norman 1988). Task scenarios are used to guide the evaluation. For the "novice use" perspective, evaluators think about novice users and answer questions as to whether a novice user can successfully go through each task steps. For the "expert use", evaluators think about expert users and notice issues such as short-cuts, appearances, information organizations as they perform tasks. For the "error handling" perspective, evaluators derive possible error types and check the interface against these error types to see how well it minimizes errors and assists error recovery. The perspective-based method gives more structure to the HE by assigning evaluators different responsibilities using perspectives, and stipulate evaluation procedures within each perspective. The evaluation result of two web-based interfaces shows that the perspective-based method can find more usability problems than the HE method.

The **metaphor of thinking** (MOT) method (Hornbæk, Frøøkjær 2004) didn't explicitly say that it is an extension of HE method, but its strong similarity with the HE procedure well qualify it as an extension. It was developed based on the classical introspective psychology. It aims to focus inspection on users' mental activity by incorporating five metaphors of human thinking: habit formation, stream of thought; awareness and association, the relationship between utterances and thought, and knowing. During the evaluation, evaluators first get familiarized with the application, and then they try to finish three tasks and meanwhile take the perspective of each of the metaphors to find usability problems in two passes. They can think of additional tasks and continue to find problems with the help of the metaphors if time allows. While HE provides simple guidelines to encourage straightforward interpretation, MOT provide complex guidelines and require evaluator's active interpretation. Experimental results show that MOT discovers more severe usability problems that are more complex to fix than HE. It also takes less time to conduct the MOT evaluation.

4.3 Extension method with a conformance rating scale

A method that extended heuristic evaluation with a conformance rating scale has been used in some studies (Mariage, Vanderdonckt 2000; Avouris et al. 2001; Sutcliffe 2001; Agarwal, Venkatech 2002; Berry 2003; McGrow et al. 2004). Evaluators are presented with a set of heuristics and requested to rate the interface based on degree of conformance to each heuristic with a rating scale. Sometimes, evaluators were also instructed to write down the rationale for their rating (Sutcliffe 2001) or suggestions for design (Avouris et al 2001) which can be the basis for finding and fixing usability problems The rating scale can be dichotomous, 5-point or 10-point ranging from lowest to highest conformance level. In additional to a list of usability problems found by traditional HE, this extended form produce quantitative data on the conformance rating of the evaluated system to each heuristic. This approach provides an overall assessment of the evaluated system in terms of its weakness and strength, and helps pinpoint the area of problems with the system to direct further corrective efforts. It can also give a quality index of the system which helps choosing among competing design options. For example, Agarwal and Venkatech (2002) developed an evaluative instrument to rate the usability of firm websites of different industries based on Microsoft usability guideline and this type of heuristic evaluation. Evaluators assume the role of a consumer or an investor when assessing usability. They first assign weight to each of five usability categories and distribute the weight over each subcategories. Then they rate website in terms of its quality regarding each subcategory. A weighted rating for each subcategory can be calculated to get the overall rating of the website.

5. Tool support for heuristic evaluation

Software tools and instruments have been developed to assist HE process to improve the efficiency of evaluators and achieve more rigorous results. IBM developed a Heuristic Evaluation Database (HEDB) to support the collaborative work of the evaluation manager and the evaluators. HEDB implements the key tasks of entering findings, assigning severity to problems, and editing duplicates in the HE process (Paddison, Englefield 2003). It has been well received by practitioners.

Since HE can be described as a creative brainstorming process, it might benefit from anonymous, parallel production with the help of collaborative software. Lowry and Roberts (2003) examined ways to increase the productivity of HE by using collaborative system. Experiment result shows that, with the collaborative software, evaluators are aware of usability problems found by others. So duplicate work is avoided, and a consensus of aggregate problem set can be found much faster.

Another instrument that can improve the HE result is the structured report format for usability problems (Cockton et al. 2003). The use of the report format can change the evaluator's behavior by demanding more reflection, and in turn result in fewer false positives and more appropriate heuristic usage.

6. Reappraisal

The free-formed structure of HE has made it a widely used discount method. But it also leads to many of its drawbacks (Cockton, Woolrych 2002). HE method has left plenty of room for improvement. Much research has been conducted to improve HE method.

A set of heuristics is intended as mnemonic framework that can cue the deeper knowledge body held by an evaluator defined by the guidelines and existing expertise (Paddison, Englefield 2003). But for novice evaluators, the usability related knowledge body is not well developed. Therefore, the sole resource for their evaluation is the heuristic set, which is not enough in many cases. They may pick user tasks and system features randomly to evaluate. Literatures have indicated four ways to give more structure to the HE method and better orient the evaluators in the search space. When the evaluation is more structured, the evaluators will feel better guided.

(1) Provide structure with modified heuristics set

Some heuristic set was explicitly divided into several parts. For example, heuristic set for websites design was divided into five parts: information design, consistency, navigation, operation, and errors (Rau, Liang 2003). The latent structure in the heuristic set might help evaluators to form more structured approach during evaluation. They may focus on detecting usability problems from each usability aspect at a time. The sHEM (Kurosu et al. 1997, 1999) method is a good example for providing structured heuristics. By dividing evaluation into sub-session, the method enables the evaluators to focus on issues in one usability sub-category in each sub-session at a time. Therefore, a larger body of usability aspects can be considered in a non-threatening way.

(2) Provide structure with problem areas

This approach tells the evaluator the most important problem areas to look at. For example, the HE-plus method provides evaluators with a usability problem profile containing problem areas (Chattratichart, Brodie 2002, 2004). This makes sure that evaluators will examine the important problem areas. Hence, comprehensive evaluation results can be produced.

(3) Provide structure with evaluation procedure

This approach tells the evaluators specifically what to do during the evaluation. For example, the perspective-based method has stipulated exact tasks to do and questions to answer with each perspective. The abstract tasks used in SUE (Matera et al. 2002; Angeli et al. 2003) also provide very stringent structure by stipulating the system elements to examine, and the related questions to answer regarding each elements.

(4) Provide structure with tasks

Six out of the seven extended HE methods listed in table 2 provide typical task scenarios to evaluators. For example, the heuristic walkthrough method introduced a task-based evaluation pass to make sure elements related to important tasks are evaluated (Sears 1997). Tasks give the evaluators better focus and understanding of the system, which, in turn, improve the evaluation result.

The heuristic evaluation is a subjective process (Doubleday et al. 1997). In addition to the checklist of general heuristics to be considered, the evaluator is also allowed to consider any additional usability principles or results that come to mind which may be relevant for any specific interface element (Nielsen 1993). Thus, the HE method sometimes serves as a method for inspiring evaluators (Cockton, Lavery, Woolrych 2003). But the inspiration that the evaluators get from the HE method is a bit limited. Another way that has proved to be able to improve the evaluation result is to inspire evaluators with different perspectives. This is like asking the evaluators to put themselves into the shoes of other type of users. It is believed that this approach can enlarge the discovery scope by providing incentives for the evaluators to think actively and broadly. It tries to fully tap the cognitive capacity of the evaluators by keeping their eyes wide open and mind actively running during the evaluation. This approach has been used by many HE extension methods. In two evaluative subsessions of the sHEM method (Kurosu et al. 1997, 1998), evaluators are asked to think whether different users (novice, expert and users with special care) will have problems using the system. Similarly, two perspectives used in the perspective based heuristic evaluation (Zhang et al. 1999) include the novice use and expert use. Evaluators think from the perspective of novice and expert users and follow the evaluative procedure to detect possible usability problems for them. In the PAVE method (Desurvire, Thomas 1993), the evaluators need to think from ten different perspectives while finding usability problems, from the most conventional views like human factor expert to most unusual view of a worried mother and a spoiled child. The metaphor of thinking method (Hornbæk, Frøøkjær 2004) also try to guide the evaluator's mental thinking with five metaphors.

In the HE method, the evaluators inspect every interface element against a list of heuristics. The result of applying the HE method is a system conforming to all of the heuristics if all identified problems are fixed. Therefore, the quality of the heuristics is directly related to the quality of evaluation results. If the heuristics doesn't cover all important aspects of the computing system usability, then the resultant system may not be good enough even if the evaluation went well. The list of heuristics should be developed carefully. Nielsen and Molich (1990) recommended that a good set of heuristics should be small (e.g., around 10) in number, so that the inspectors could have an easy time remembering, and being reminded of the heuristics while they are detecting usability problem. Nielsen's heuristic set has a manageable size of 10. It was speculated that Nielsen limit the number of heuristics to fit the limitation of human memory with the chunk size of 7 plus or minus 2 (Kurosu et al. 1997). Though larger number of heuristics may address more usability problems, they may not be as easily maintained in the evaluator's working memory. Therefore, when developing domain specific heuristic set, the total number should not be too large.

Many heuristic sets listed in table 1 are still under revision. We can notice that different heuristic sets were developed for the same domain of e-learning websites (Reeves et al. 2002; Evans, Sabry 2003). The reason is that they use different theoretical basis for heuristic set development. With more application case studies and experiments, we would expect that the disparate sets can be merged into a comprehensive set. Heuristic set should be refined iteratively until a validated heuristic set emerge for each domain. Practitioners can then choose these domain specific heuristics as tools to evaluate the corresponding systems.

As the usability inspection method such as the heuristic evaluation develop and mature, we can understand better the facts that affect the quality of evaluation results, and can design better methods to achieve result with high thoroughness, validity and reliability.

7. Future research

This review study shows the plethora of research devoted to optimize the usability inspection method: HE method. With better supported evaluation procedure, and better fitted heuristic set, the HE and its extensions should be more ready to serve as effective discount method. The future research direction related to HE method can be summarized in the following aspects:

(1) Extension methods to HE need to be applied to various computing systems to test their applicability and generalizability.

(2) More research is needed to compare the effectiveness among the different extension method.

(3) Get better understanding of the differences between the extended methods and the traditional HE method in terms of evaluator's cognition process.

(4) Study the difference between the expert and novice evaluators while using the HE method and extended HE methods.

(5) Usability inspection methods need to be applied to more domains to improve their usability. More domain-specific heuristics need be developed and refined to help give precise and relevant evaluation results.

(6) More extended method can be developed to further improve the HE method. A typical HE procedure includes four phases: pre-evaluation training, the actual evaluation, debriefing session, and a severity rating phase (Nielsen 1994b). Currently, most extended methods aimed to improve the actual evaluation phase of the HE method. More improvement may be achieved by improving more or all phases of the HE method.

(7) More software tools can be developed to assist the heuristic evaluation process.

(8) With the presence of data from many empirical studies, some kind of meta-analysis can be performed to better determine the factors that play important roles in detecting usability problems.

References

Agarwal R, Venkatesh V, 2002, Assessing a firm's web presence: a heuristic evaluation procedure for the measurement of usability. *Information Systems Research*, 13, 168–186.

Angeli AD, Matera M, Costabile MF, Garzotto F, Paolini P, 2003, On the advantages of a systematic inspection for evaluating hypermedia usability. *International Journal of Human-Computer Interaction*, 15, 3, 315–335.

Avouris NM., Tselios N, Tatakis EC, 2001, Development and evaluation of a computer-based laboratory teaching tool. *Computer Applications in Engineering Education, 9*, 1, 8–19.

Baker K, Greenberg S, Gutwin C, 2001, Heuristic evaluation of groupware based on the mechanics of collaboration. *Proceedings of the 8th IFIP Working conference on Engineering for Human-Computer Interaction*, May 11–13, Toronto, Canada.

Baker K, Greenberg S, Gutwin C, 2002, Empirical development of a heuristic evaluation methodology for shared workspace groupware. *Proceedings of the 2002 ACM conference on computer supported cooperative work*, November 16–20, New Orleans, Louisiana.

Bendoly L, Goldman KH, 2003, Investigating heuristic evaluation: a case study. *Proceedings of the Museums and the Web 2003.* March 19–22, Charlotte, North Carolina.

Berry B, 2003, Adapting heuristics for notification systems. *Proceedings of 41st Annual ACM Southeast Conference*, Savannah, GA, 144–149.

Blandford A, Keith S, Connell I, Edwards H, 2004, Analytical usability evaluation of digital libraries: a case study. *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, June 7–11, Tuscon, Arizona, USA.

Chattratichart J, Brodie J, 2002, Extending the heuristic evaluation method through contextualization. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*, September 30–October 4, Baltimore, MD.

Chattratichart J, Brodie J, 2004, Applying user testing data to UEM performance metrics. *Proceedings of the CHI 2004 Conference on Human Factors in Computing Systems*, April 24-29, Vienna, Austria, 1119–1122.

Cockton G, Woolrych A, 2001, Understanding inspection methods: lessons from an assessment of heuristic evaluation. [in:] Blandford AJ, Vanderdonckt J (eds.) *People & Computers XV*, Springer-Verlag, 171–192.

Cockton G, Woolrych A, 2002, Sales must end: should discount methods be cleared off HCI's shelves? *Interactions*, September + October, 13–18.

Cockton G, Lavery D, Woolrych A, 2003, Inspection-based evaluation. [in:] Jacko JA and Sears A (eds.) *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1118–1138.

Cockton G, Woolrych A, Hall L, Hindmarch M, 2003, Changing analyst's tunes: the surprising impact of a new instrument for usability inspection method assessment. [in:] Staphanidis C and Jacko J (eds.) *Proceedings of HCI 03 Conference*, June 22–27, Crete, Lawrence Erlbaum Associates, 145–161.

Connell IW, Hammond NV, 1999, Comparing usability evaluation principles with heuristics: problem instances vs. problem types. [in:] Sasse MA and Johnson C (eds.) *Proceedings of the IFIP INTERACT '99 Conference on Human-Computer Interaction*, Amsterdam: IOS, 621–629.

Desurvire H, Caplan M, Toth JA, 2004, Using heuristics to evaluate the playability of games. *Proceedings of the CHI 2004 Conference on Human Factors in Computing Systems*, April 24–29, Vienna, Austria.

Desurvire H, Thomas JC, 1993, Enhancing the performance of interface evaluators using non-empirical usability methods. *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society*, October11–15, Seattle, Washington.

Doubleday A, Ryan A, Springett M, Sutcliffe A, 1997, A comparison of usability techniques for evaluating design. *Proceedings of Designing Interactive Systems (DIS'97): Processes, Practices, Methods, and Techniques*, Amsterdam, the Netherlands, ACM Press, 101–110.

Drury J, 2001, Developing heuristics for synchronous collaborative systems. *Proceedings of the CHI 2001 Conference on Human Factors in Computing Systems*. March 31–April 5, Seattle, Washington.

Evans C, Sabry K, 2003, Evaluating of the interactivity of web-based learning systems: principles and process. *Innovations in Education and Teaching International*, 40, 1, 89–99.

Fakun D, Greenough RM, 2002, User-interface design heuristics for developing usable industrial hypermedia applications. *Human Factors and Ergonomics in Manufacturing*, 12, 2, 127–149.

Fernandes G, Holmes C, 2002, Applying HCI to music related hardware. *Proceedings of the CHI 2002 Conference on Human Factors in Computing Systems*, April 20–25, Minneapolis, Minnesota.

Fu L, Salvendy G, Turley L, 2002, Effectiveness of user testing and heuristic evaluation as a function of performance classification. *Behaviour & Information Technology, 21*, 2, 137–143.

Garzotto F, Mainetti L, Paolini P, 1995, Hypermedia design, analysis, and evaluation issues. *Communication of ACM*, 38, 74–86.

Garzotto F, Matera M, Paolini P, 1998, Model-based heuristic evaluation of hypermedia usability. *Proceedings of the Working Conference on Advanced Visual Interfaces*, May 24–27, L'Aquila, Italy.

Greenberg S, Fitzpatrick G, Gutwin C, Kaplan S, 1999, Adapting the Locales framework for heuristic evaluation of groupware. Proceedings of OZCHI'99, 28–30.

Gutwin CG, 2000, The mechanics of collaboration: developing low cost usability evaluation methods for shared workspaces. *Proceedings of the IEEE 9th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*. March 14–16, 2000. Gaithersburg, Maryland.

Hertzum M, Jacobsen NE, 2001, The evaluator effect: a chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction,* 13, 4, 421–443.

Hornbæk K, Frøøkjær E, 2004, Usability inspection by metaphors of human thinking compared to heuristic evaluation. *International Journal of Human-Computer Interaction,* 17, 3, 357–374.

Jeffries R, Wharton C, Uyeda KM, 1991, User Interface evaluation in the real world: a comparison of four techniques. *Proceedings of CHI'91 Conference on Human Factors in Computing Systems*, April 27–May 2, New Orleans, LA.

Johnson D, Wiles J, 2003, Effective affective user interface design in games. *Ergonomics*, 46, 13-14, 1332–1345.

Kamper RJ, 2002, Extending the usability of heuristics for design and evaluation: lead, follow, and get out of the way. *International Journal of Human Computer Interaction*, 14, 3-4, 447–462.

Kline RS, Javahery H, Donayee M, Rilling J, 2002, Quantifying developer experiences via heuristic and psychometric evaluation. *Proceedings of the IEEE 2002 Symposia on Human Centric Computing Languages and Environments*, September 03–06, Arlington, Virginia

Köykkä M, Ollikainen R, Ranta-aho M, 1999, Usability heuristic guidelines for 3D multi-user worlds. *Proceedings of the 1999 Conference on Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia*, November 28–30, Wagga Wagga, Australia, 52–57.

Kurosu M, Matsuura S, Sugizaki M, 1997, Categorical inspection method-structured heuristic evaluation (sHE). *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics "Computational Cybernetics and Simulation",* October 12–15, Orlando, Florida.

Kurosu M, Sugizaki M, Matsuura S, 1998, Structured heuristic evaluation (sHEM): A more productive method to find out usability problems. *Proceedings of the Usability Professionals' Association Conference,* June22–26, Washington DC.

Kurosu M, Sugizaki M, Matsuura S, 1999, A comparative study of sHEM (structured heuristic evaluation method). [in:] Bullinger HJ, Ziegler J (eds.) Human-computer interaction: Ergonomics and user interfaces. *Proceedings of HCI International 99*, Lawrence Erbaum Associates, Mahwah, NJ, 1, 938–942.

Lathan CE, Sebrechts MM, Newman DJ, Doarn CR, 1999, Heuristic evaluation a Web-based interface for Internet Telemedicine. *Telemedicine Journal*, 5, 2, 177–185.

Law LC, Hvannberg ET, 2002, Complementarity and convergence of heuristic evaluation and usability test: a case study of UNIVERSAL brokerage platform. *Proceedings of the 2nd Nordic Conference on Human-Computer Interaction*, October 19–23, Århus, Denmark.

Levi MD, Conrad, FG, 1996, A heuristic evaluation of a world wide web prototype. *Interactions*, July+August, 51–60.

Lewis J, 2006, Usability testing. [in:] Salvendy G. (eds.) *Handbook of Human Factors and Ergonomics*, 3rd Edition, John Wiley & Sons, New York, NY.

Lowry PB, Roberts TL, 2003, Improving the usability evaluation technique, heuristic evaluation, through the use of collaborative software. *Proceedings of the 9th Annual Americas Conference on Information Systems (AMCIS)*, August 4–5, Tampa, 2203–2211.

Mack RL, Montaniz F, 1994, Observing, predicting, and analyzing usability problems. [in:] Nielsen J and Mack RL (eds.) *Usability Inspection Methods*. John Wiley & Sons, New York, NY, 295–339.

Mack RL, Nielsen J, 1994, Executive summary. [in:] Nielsen J and Mack RL (eds.) *Usability Inspection Methods*. John Wiley & Sons, New York, NY, 1–23.

Mankoff J, Dey AK, Hsieh G, Kientz J, Lederer S, Ames M, 2003, Heuristic evaluation of ambient displays. *Proceedings of the CHI 2003 Conference on Human Factors in Computing Systems,* April 5–10, Ft. Lauderdale, Florida.

Mariage C, Vanderdonckt J, 2000, A Comparative usability study of electronic newspapers. P*roceedings of International Workshop on Tools for Working with Guidelines,* October 7-8, Biarritz, Springer-Verlag, London, 325–337.

Matera M, Costabile MF, Garzotto F, Paolini P, 2002, SUE inspection: An effective method for systematic usability evaluation of hypermedia. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 32, 1, 93–103.

McGrow K, Brennan AH, Preece J, 2004, Development of a tool for heuristic evaluation of healthcare information systems. *Computers, Informatics, Nursing, Journal of Hospice & Palliative Nursing*, in press.

Miller JR, Jeffries R, 1992, Interface-usability evaluation: science of trade-offs. *IEEE Software*, 9, 5, 97–98, 102.

Mills S, 1995, Usability problems of acoustical fishing displays. *Displays*, 16, 3, 115–121.

Molich R, Nielsen J, 1990, Improving human-computer dialog. *Communications of the ACM*, 33, 3, 338–348.

Muller MJ, Matheson L, Page C, Gallup R, 1998, Participatory heuristic evaluation. *Interactions*, September+October, 13–18.

Nielsen J, 1992, Finding usability problems through heuristic evaluation. *Proceedings of CHI'92 Conference on Human Factors in Computing Systems*, May 3–7, Monterey, California, 373–380.

Nielsen J, 1993, *Usability Engineering*. Boston: Academic Press.

Nielsen J, 1994a, Enhancing the explanatory power of usability heuristics. *Proceedings of the CHI'94 Conference on Human Factors in Computing Systems "Celebrating interdependence"*, April 24–28, 1994, Boston, Massachusetts.

Nielsen J, 1994b, Heuristic evaluation. [in:] Nielsen J, Mack RL (eds.) *Usability Inspection Methods*. John Wiley & Sons, New York, NY, 24–62.

Nielsen J, Landauer TK, 1993, A mathematical model of the finding of usability problems. *Proceedings of the InterCHI Conference on Human Factors in Computing Systems*, April 24–29, Amsterdam, The Netherlands.

Nielsen J, Mack RL, 1994, *Usability Inspection Methods.* John Wiley & Sons, New York, NY.

Nielsen J, Molich R, 1990, Heuristic evaluation of user interfaces. *Proceedings of CHI'90 Conference on Human Factors in Computing Systems*, April 1–5, Seattle, Washington, 249–256.

Nielsen J, Phillips VL, 1993, Estimating the relative usability of two interfaces: heuristic, formal, and empirical methods compared. *Proceedings of the InterCHI Conference on Human Factors in Computing Systems*, April 24–29, Amsterdam, The Netherlands.

Norman DA, 1988, *The Design of Everyday Things*. Basic Books, New York.

Paddison C, Englefield P, 2003, Applying heuristics to perform a rigorous accessibility inspection in a commercial context. *Proceedings of the CUU'03*, November 10–11, Vancouver, British Columbia, Canada.

Rau P, Liang SF, 2003, Internationalization and localization: evaluating and testing a website for Asian users. *Ergonomics*, 46, 1-3, 255–270.

Reeves TC, Benson L, Elliott D, Grant M, Holschuh D, Kim B, Kim H, Lauber E, & Loh S, 2002, Usability and instructional design heuristics for E-Learning evaluation, *Proceedings of World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Denver, Colorado.

Rosenbaum S, Rohn JA, Humburg J, 2000, A toolkit for strategic usability: results from workshops, panels, and surveys. [in:] Little R, Nigay L (eds.) *Proceedings of ACM CHI 2000 Conference on Human Factors in Computing Systems,* April 1–6, The Hague, The Netherlands, ACM press, New York, 337–344.

Sears A, 1997, Heuristic walkthroughs: Finding the problems without the noise. *International Journal of Human-Computer Interaction*, 9, 3, 213–234.

Shneiderman B, 1998, *Designing the user interface. Strategies for effective human-computer interaction*. (3rd Edition). Addison-Wesley, Reading, MA.

Slavkovic A, Cross K, 1999, Novice heuristic evaluations of a complex interface. *Proceedings of the CHI'99 Conference on Human Factors in Computing Systems*, May 15–20, Pittsburgh, Pennsylvania.

Somervell J, Wahid S, McCrickard DS, 2003, Usability heuristics for large screen information exhibits. *Proceedings of the Ninth IFIP TC13 International Conference on Human Computer Interaction*, Zurich, Switzerland, 904–907.

Squires D, Preece J, 1999, Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers*, 11, 467–483.

Sutcliffe A, 2001, Heuristic evaluation of website attractiveness and usability. *Proceedings of the 8th International Workshop on Interactive Systems: Design, Specification, and Verification*, Springer-Verlag, London, UK, 183–198

Vetere F, Howard S, Pedell S, Balbo S, 2003, Walking through mobile use: Novice heuristic and their application. *Proceedings of OzCHI2003: New directions in interaction, information environments, media and technology*, November 26–28, Brisbane, Australia.

Warren P, 2004, Learning to program: spreadsheets, scripting and HCI. *Proceedings of the Sixth Australasian Computing Education Conference*, Dunedin, New Zealand.

Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T, 2003, Using usability heuristics to evaluate patient safety of medical devices. *Journal of Biomedical Informatics*, 36, 1/2, 23–30.

Zhang Z, Basili V, Shneiderman B, 1999, Perspective-based usability inspection: An empirical validation of efficacy. *Empirical Software Engineering*, 4, 1, 43–69.