Remote Usability Evaluation Methods and Tools: A Survey

Fidas Christos, Katsanos Christos, Papachristos Eleftherios, Tselios Nikolaos, Avouris Nikolaos

Human Computer Interaction Group,
Electrical and Computer Engineering Dept.,University of Patras, Rio, Greece
{fidas, ckatsanos, epap, nitse, avouris}@ece.upatras.gr

Abstract

Effective and efficient usability evaluation methodologies are required in order to develop and maintain software applications that meet user requirements and expectations. This is particularly relevant to applications that change frequently during their lifecycle. Remote usability evaluation provides such a solution. Numerous approaches and techniques have been proposed to allow remote usability evaluation of software applications, and the purpose of this paper is to categorize and evaluate these approaches. We also discuss relevant issues such as user data collection and analysis. After surveying the proposed approaches and identifying pros and cons, we conclude with promising directions for future research in this area.

Keywords: Human Computer Interaction, Usability Evaluation, Remote Usability Evaluation Methods, Remote Usability Evaluation Tools

1. Introduction

The rapidly expanding research in software usability evaluation is based on the premise that usable software products increase user satisfaction, effectiveness and efficiency, improve product quality while simultaneously reduce maintenance and support costs [Nielsen (1993), Dix et Al. (1998)]. Usability refers to whether a system can be used with effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in a particular context of use [ISO 9241-11].

Methodologies for software development need to take advantage of user centred design approaches which give extensive attention to the needs and limitations of the users at each stage of the design process [Vredenburg et. Al.(2002)]. Furthermore, a relatively recent trend in web applications which is related to the *rich internet application frameworks*, further enrich the interactivity of web applications bridging the gap between web-based and stand alone desktop applications. With these new technologies, development times are rapid and changes occur frequently, often without a chance to re-evaluate the entire web site [Baresi et. Al. (2000)]. Within this

context, where the potential audience for a web system is mainly geographically dispersed or encompasses a wide range of demographic groups and new software development frameworks change the user interaction model with an application, traditional in-lab usability approaches tend to be less relevant and not cost effective.

Remote usability evaluation, which is characterized by the separation in space and/or time among the evaluators and the respondents, is an emerging field within Human Computer Interaction which seems to offer a promising solution to the aforementioned problems [Krauss (2003), Hartson et. Al. (1996), Ivory et. Al. (2001), Tullis et. Al. (2002), West & Lehman (2006)].

With an aim to give a comprehensive and critical survey of current remote usability evaluation methods and tools, this paper is organized as follows: In Section 2, we give a detailed definition of remote usability evaluation and the conceptual models behind its main categories. A review of available remote usability evaluation methods and evaluation criteria for classifying widely used remote usability tools, are discussed in Section 3. We conclude this paper with a discussion of several promising directions for research in the area of remote usability evaluation in Section 4.

2. Remote Usability Evaluation Approaches

In this section, we review existing approaches to perform remote usability activities. After providing a detailed definition of remote usability evaluation we classify remote usability evaluation methods into two major categories and we discuss the motivation and general approach of each category.

2.1 Remote Usability Evaluation Definition

Remote usability evaluation can be considered as a new paradigm for evaluating an interactive system in a cost effective, fast and efficient manner. Remote usability evaluation combines empirical and analytic usability evaluation techniques that are usually performed separately. In general, the differences between in-lab usability and remote usability approaches are: a) the usability engineer and the respondents are at different locations, b) the usability engineer is obliged to use a software tool for observing and analyzing the respondents interactions with the software system and c) the usability engineer and respondents communication is usually mediated by a software tool.

Benefits connected with remote usability activities are: a) it makes easier to reach respondents in diverse geographic areas, b) it is useful for evaluating systems that have been designed for hard-to-reach, decentralized groups of users for whom it would be difficult to schedule interviews at a single location, c) it is cost effective since it reduces travel costs for the usability engineers and the respondents, d) a larger and diverse pool of participants is accessible, thus a worldwide audience can be

reached, e) participants can use the product in a more realistic environment and f) it involves more participants thus providing more convincing usability results.

Limitations or drawbacks in performing remote usability activities are mainly related with the additional "degree of separation" among the usability engineer and the respondents including: a) difficulties to build mutual understanding and trust due to limited communication channels, b) difficulties to capture respondents' facial expressions and other nonverbal cues, c) limited communication channels dictated by the functionalities of the mediating software tools and d) the social and cultural context of an international pool of users may sometimes bias the results.

In general two categories of remote usability techniques can be distinguished: a) moderated and b) automated techniques. These two categories are classified according to the co-presence of the usability engineer and the respondents, the communication channels they use and the usability methodologies they support. Below, we discuss the motivation and general approach for each category.

2.2 Moderated Remote Usability Evaluation

Moderated remote usability evaluation approaches are characterized by the geographical separation between the usability engineer and the respondents. Moderated remote usability activities are similar with in-lab testing but with a level of separation between the usability expert and the participants. Thus, the main challenge in moderated remote usability is to overcome this level of separation. In this context, communication among the involved participants includes two main spaces of interaction: (a) the discourse space where interaction occurs via audio, text or videoconferencing among the involved participants and (b) the task space where interaction occurs through a shared workspace [Dix et. Al. (1998)].

From the evaluator's point of view, the challenges associated with moderated remote usability evaluation are: a) establishing common concepts with the remote participants via synchronous communication, b) observing participants' workstations, c) interfering in cases where participants run into difficulty while interacting with the evaluated software system, d) recording both the conversation and the participant's screen and e) analyzing the recorded data (Figure 1).

Since, moderated remote usability activities require the installation of a variety of services to support the evaluation process, the main challenge from the participant's point of view is to achieve a high level of transparency related with the tools required to support and mediate the interaction with the evaluators. Thus, in order to perform successfully moderated remote usability studies, easy to use and install services for the participant's side are important.

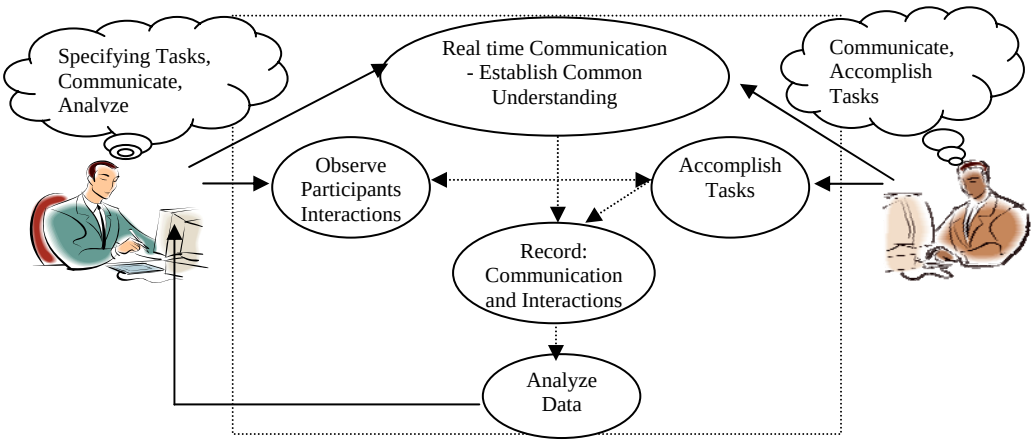

Figure 1. *Conceptual Model of Moderated Remote Usability Activities*

The usability engineer communicates in real time with the participant aiming to the establishment of a shared understanding. The communication channels vary according to the used software tool and can be audio, video or text based. Commonly, computer-based video conferencing tools display live video in a window of the person or persons that participate in the evaluation process. Visual cues such as gestures and facial expressions help to establish rapport between conference participants engaging discussions. Another important requirement for moderated remote usability activities is the real time observation of the participants' screen. This is achieved with tools that support real-time sharing of application windows between two or more workstations. Once sharing has occurred, both the sender and the receiver can view the shared application or window as if it was actually running on each of their respective machines.

2.3 Automated Remote Usability Evaluation

Automated remote usability evaluation does not require the usability engineer and the respondents to be in the same place or time. Conceptually, automated remote usability evaluation methods combine mainly two types of evaluation techniques that are usually applied separately: empirical testing and model-based evaluation. Empirical testing is a technique according to which information gathered for the usability evaluation is derived from actual users of the system engaged in real tasks.

In this context, automated remote usability techniques can be either user-reported oriented or automated usage tracking. In the first case the participants report their own behaviour through a browser window answering survey questions or reporting critical incidents in their interaction (user-reported critical incident method [Castillo et. Al. 1998]). Usually, as the users perform specific tasks and navigate through a web site, they enter their feedback or answer task-specific questions in the browser frame.

In the second case, behavioural usage data is automatically collected by the system the user interacts with, to be either manually or automatically analyzed afterwards.

As it can be seen in Figure 2 automated remote usability activities require from the usability engineer to describe in detail the exact sequence of actions a user will perform, which is usually achieved through detailed task analysis. Furthermore, in this context, model-based evaluation approaches are commonly used in order to predict certain aspects of user performance such as task completion times or difficulty of learning a task sequence. The most common model-based approach to estimating usability is the GOMS method of [Card et. Al.(1983)], used to describe an ideal error-free behavior. From this step the designer's model is created which is then compared with the various user mental models build from user interaction data derived from log files [Paganelli et. Al (2003)]. Since, automated remote usability activities are performed without the direct observation of the user interactions, it is important to save detailed information that can be afterwards analyzed from the usability engineer.

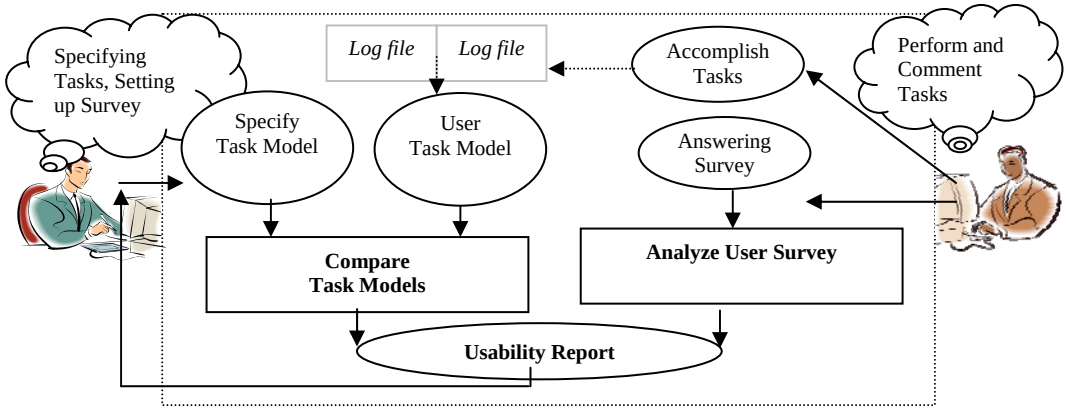

Figure 2. *Conceptual Model of Automated Remote Usability Activities*

Automated remote usability systems create automatically usability reports in order to highlight usability flaws. The evaluation that is performed consists of analyzing sequences of actions and finding relationships and differences among the evaluator's and the participants' interaction models, but also among the participants' models themselves.

3. Tools Supporting Remote Usability Evaluation Activities

Remote usability evaluation can implement almost every known traditional usability evaluation method. Aiming to identify a taxonomy for remote (moderated and automated) usability evaluation tools we provide an overview regarding the features required for performing these usability evaluation methods. Table 1 presents the requirements that each of the traditional usability evaluation methods poses in a remote context. We have classified the widely used usability evaluation methods in

the following four categories: a) Usability Inspection, b) Usability Testing, c) Exploratory and d) Analytic methods.

Table 1. *Applying traditional usability evaluation methods in a remote context*

Usability evaluation Methods		**Remote Usability Requirements**	
		Minimum Requirements	*Improving Features*
Usability Inspection	Heuristic Evaluation / Cognitive Walkthrough/ Guidelines Application	Direct Communication(Text), Common Repository (file sharing)	Direct Communication (Audio, Video), Annotation Tools (Shared Whiteboard for Inspection Annotation)
Usability Testing	Performance Measurement	Data Gathering (UI events Logging)	Data Analysis Tools (Summary, Statistics, Visualization)
	Think Aloud Protocol / Question-Asking Protocol	Direct Communication(Audio),Application Sharing	Direct Communication (Video), Data Gathering (Video, Audio, Screen), Post-Hoc Discussion (Shared Whiteboard), Multiple Observers' support, Participant recruitment
Exploratory	Surveys	On-line Questionnaire	Questionnaire Analysis (summaries, statistics), Participant recruitment
	Interviews	Direct Communication(Text)	Direct Communication(Audio, Video), Data Gathering (Video, Audio), Participant recruitment
	Focus Groups	Direct Communication(Text)	Direct Communication(Audio, Video), Application Sharing, Data Gathering (Video, Audio)
Analytic	Task Analysis	Task Definition, Data Gathering (UI events Logging)	Data Analysis (Reconstructing user interactions, Task model comparison), Participant recruitment

It is important to mention that in the analytic, exploratory and user testing methodologies two distinct groups of users are involved: a) the usability experts and b) the respondents, whereas in the inspection methods only usability experts are involved. For the latter, requirements are identified for performing these methods by experts located in different areas. In this context, the minimum requirements for applying inspection methods are to allow the cooperation among usability experts in terms of defining the details of the evaluation (e.g. guidelines) and communicating results through a common repository. An improvement to the remote usability inspection approach is the existence of an annotation tool that enables the experts to underpin the identified usability flaws on specific areas in the system interface.

In the next two sections we classify well-known automated and moderated tools according to the usability evaluation methodologies they support and to the features they provide in every stage of a remote usability activity.

3.1 Tools Supporting Moderated Remote Usability Activities

In moderated remote usability activities the underlying idea is that the evaluator observes in real-time users performing tasks, and strives to understand their behavior in order to detect usability flaws. The activities can range from participatory design exercises and storyboard walkthroughs to formative evaluations and user testing that adopts varying research designs. In the context of geographically separated users and evaluators, information and communication technologies are required to mediate and support the moderated remote usability activities. Such technological solutions need to support all the important phases of a moderated remote usability activity, namely the *preparation*, *execution*, *data gathering* and *data analysis*.

The challenges in the *preparation* of a moderated usability study are related to the participants recruiting and to the installation transparency from the participant's point of view. For the participants recruiting two are the dominant approaches in the available tools: embedded email invitation and live-recruiting. Live-recruiting deploys an online screener to intercept people in the middle of their real-life tasks, and watch them live in their native task environment. Thus, live-recruiting requires no scheduling in advance but assumes either that the evaluated system is a web application or that there is a web site associated with it. The main issues regarding the *execution* of moderated remote usability activities are related to the establishment of a shared discourse and task space, observers' support and other technical details such as platform interoperability and network connectivity. In this context, computer mediated communication and collaboration tools are used to allow participants to interact with the system under review in real time. These tools provide useful functionalities - such as application sharing and audio/video conferencing – that allow the evaluator and the participants to establish a common understanding and achieve collaboration. The main issues that need to be addressed when considering the *data gathering* and *data analysis* phases relate to the type of the gathered data and the techniques offered to analyze them, identify and report possible usability flaws in the user interaction.

Table 2 presents a classification of the most known and widely used available tools for moderated remote usability activities based on the aforementioned phases. At this point it should be mentioned that there is a wide variety of other available communication and collaboration tools – such as GoToMeeting, Co-Pilot, Microsoft NetMeeting, Skype with extensions, Microsoft Windows Messenger, ShowMe and Lotus Sametime – that could be also used to meet some of the requirements of a moderated remote usability activity.

Table 2. – Taxonomy of Tools Supporting Moderated Remote Usability Activities

			UserVue	Ethnio	WebEx	Adobe Connect+
RESEARCH DESIGN		Performance Measurement	Yes	Yes	No	No
		Think Aloud	Yes	Yes	Yes	Yes
		Question Asking	Yes	Yes	Yes	Yes
		Interviews	Yes	No	Yes	Yes
		Focus Groups	Yes	No	Yes	Yes
		Task Analysis	Yes	No	Yes	Yes
PREPARATION		Participants Recruiting Support	Email Invitation	Live-Recruiting	Email Invitation	No
		Client Install Required	Client Toolbar	Browser Plug-In	Browser Plug-In	Flash Plug-In
EXECUTION	Shared Space	Application Sharing	Yes	Yes	Yes	Yes
		Shared Whiteboard	No	No	Yes	Yes
		Remote Control	No	No	Yes	Yes
	CMC#	Instant Messaging	Yes	No	Yes	Yes
		Audio Conferencing	Tele-conference	Tele-conference	Tele-conference or VoIP	Tele-conference or VoIP
		Video Conferencing	No	No	Yes	Yes
	Observer Support	Multiple Observers	10	10	15	10
		Communication Mechanisms	Chat	No	Chat	Chat
	Technical Issues	Firewall Problems	No	Yes	No	No
		Platform Compatibility	Windows, IE, Firefox	Windows, IE	All	All
		Security	SSL	SSL	SSL, AES	SSL
DATA GATHERING		Chat Messages	Yes	No	Yes	Yes
		UI Events	Yes	Yes*	No	No
		Audio	Yes	Yes	Yes	Yes
		Screen Video	Yes	Yes	Yes	Yes
		Participant Video	No	No	Yes	Yes
DATA ANALYSIS		Data Sources Sync	Yes	Yes	Yes	Yes
		Annotation Tools	Yes	No	Yes	Yes
		Summative Statistics	No	No	No	No
		Data Visualization	No	No	No	No

* Applies only for web applications #Computer Mediated Communication +Former Breeze

A final point to be made is that combinations of different software applications can be also used for the same purposes. One typical example is the case where screen video

recording tools (e.g. Camtasia, Captivate) and web conferencing tools (usually WebEx or GoToMeeting) are combined in a "reverse-setup" allowing the participant to control the evaluator's computer where the evaluated product is installed. The advantage of this approach is that it can exploit some advanced data gathering features offered by such screen video recording tools to enrich the data gathered (e.g. log UI events) but the main drawback is that the participant's interaction experience can be greatly diminished by network delays.

3.2 Tools Supporting Automated Remote Usability Activities

Contrary to moderated methods, in automated remote usability settings the usability experts do not observe respondents in real time. We distinguish two alternative approaches - which can be also combined - of gathering usage interaction data: a) user reporting techniques for gathering subjective feedback and b) usage tracking techniques which provide valuable insight regarding user behavioral patterns by analyzing clickstream data [Fidas et. Al (2006)]. Commonly, the results are automatically reported containing information about aggregate verbatim answers, completion time, abandon frequency, navigation paths and satisfaction indicators.

Participants' recruiting is more challenging than in moderated settings due to the large amount of users required in order to infer valid conclusions. Due to this fact the contribution of panel acquisition services is often required. However, other options such us email invitation and live-recruiting are also available. In some activities, recruitment isn't needed at all, especially in web applications evaluation, where conclusions can be drawn by analyzing log-files using data mining techniques or by reconstructing user interactions. In some cases users are required to install some browser plug-in in order to be able to participate in the study. This installation is usually a minor inconvenience since most applications are using well known plug-ins or the installation has no demanding system requirements and firewall complications.

The *data gathering* issue is of particular interest in existing automated tools because of the many different research designs. Furthermore, the *data analysis and presentation* techniques supported are of particular importance due to the large amount of raw data and the need to have several levels of abstractions.

In Table 3 a classification of the most used available tools for automated remote usability activities is presented. MindCanvas is more a participatory design exercise tool but has also the capability of conducting formal survey questions. Furthermore, it is worthy to mention that there are also some tools more appropriate for specific purposes - like on line survey tools (Zoomerang and Survey Monkey) or online longitudinal qualitative studies and blog-based diaries (KDA Relevations).

Table 3. – Taxonomy of Tools Supporting Automated Remote Usability Activities

		MindCanvas	**UserZoom**	**ClickTale**	**Relevantview**
RESEARCH DESIGN	Heuristic Evaluation	No	No	No	No
	Cognitive Walkthrough	No	No	No	No
	Guidelines Application	No	No	No	No
	Performance Measurement	No	Yes	Yes	Yes
	Surveys	Yes	Yes	No	Yes
	Task Analysis	No	No	No	No
PREPARATION	Participants Recruiting Support	P.A.S	P.A.S, Invitation	Real users	P.A.S, Invitation
	Task Definition	Yes	Yes	No	Yes
	Client Install Required	Macromedia Flash plug	Browser Plug-In	No	No
EXECUTION	Firewall Problems	No	No	No	
	Platform Compatibility	All	IE6.0+, FireFox1.0+	IE6.0+, FireFox1.0+	All
DATA GATHERING	Subjective data	Questionnaires, Card Sorting	Questionnaires	No	Yes
	Actions sequences	Yes	Yes	Yes	Yes
	Task completion	Yes	Yes	No	Yes
	Time on task	No	Yes	Yes	Yes
	Mouse Clicks	No	Yes	Yes	Yes
	Error rates	No	Yes	No	Yes
DATA ANALYSIS	Sequence Analysis	No	Yes	No	Yes
	Summative Statistics	Yes	Yes	Yes	Yes
	Data Visualization	Yes	Yes	No	No

*P.A.S (Panel Acquisition Service)

4. Discussion and Conclusion

Highly interactive applications, especially when they need frequent updates and modifications, require cost effective usability methodologies that can be applied rapidly and are easy to use. Remote usability evaluation is a new paradigm that provides a solution to this challenge. We argue that remote usability evaluation techniques are cost effective and efficient approaches that can be used to evaluate both standalone and web-based applications in a complementary way to the traditional usability methodologies conducted in usability laboratories. Relevant studies [West & Lehman (2006)] suggest that remote usability evaluation tests

capture very similar usability-related information with in-lab activities. Both types of evaluation appear to capture the most significant issues in the system under review. Thus, it is clear that remote usability evaluation, implemented properly, is a valuable method to asses the user experience.

The main challenges in using moderated remote usability techniques are: a) to overcome the additional level of separation among the usability engineer and the participants and b) to succeed a certain level of transparency among the tools used for supporting the usability evaluation and the software system under evaluation. Thus, we argue that moderated remote usability requires an easy to use and to install service that will not add any effort on the side of the respondents. Automated remote usability is not just an offering of on-line questionnaires to the respondents. It is rather considered as the ability of the system to provide the necessary data to the usability engineers in order to evaluate the usability of a system by modeling user interactions, and analyzing subjective feedback from the users.

A general conclusion derived from the survey is the absence of an integrated tool to support sufficiently the phases defined in this paper for designing and implementing remote usability evaluation studies. As far as moderated remote usability activities are concerned, the combination of different tools that support the diverse phases of the design and implementation of such activities seems to be a good solution for most research designs. As an example, by combining the different types of data gathered by *UserVue* with the advanced data analysis techniques offered by *Morae*, the usability expert can gain a better insight into users' behavioral patterns. In accordance, in automated remote usability evaluation, we found that *RelevantView* seems to provide the most complete solution, if a combination of user reporting techniques and performance measurement indicators are required. On the other hand, if the interest is shifted towards participatory design techniques, then *MindCanvas* is a good choice. The transformation of semi-structured methods like card sorting to game-like elicitation methods offered by *MindCanvas* improves user engagement and makes participant recruitment easier.

Another conclusion derived is that remote usability inspection methods seem not to be supported by any of the presented tools. To some extent this is expected, since these methods are different in the problems they present and seem to benefit less from the remote character of the techniques, as presented in this paper. However, taking into consideration that usability experts can be also geographical dispersed, a tool with useful integrated functionalities (such as annotation tools, common repository and communication) could potentially enhance the overall evaluation process.

Remote usability is a challenging and interesting area of research and technology. As new technologies change the user interaction model and web and standalone applications are becoming more alike, remote usability approaches are becoming of more general value.

References

Baresi, L., Garzotto, F., Paolini, P. (2000). *From Web Sites to Web Applications: New Issues for Conceptual Modeling*, ER 2000 Workshop, LNCS 1921, pp. 89-100.

Card, S., Moran, T, Newell, A. (1983). *The psychology of human computer interaction*, Hillsdale, NJ:Lawrence Erlbaum Associates, ISBN 0-8985-9243-7.

Castillo, J., Hartson, H, Hix, D. (1998*). Remote usability evaluation: can users report their own critical incidents?* in Proc. of the SIGCHI conference on Human factors in computing systems: common ground, p.253-254, LA, California.

Dix, A., Finlay, J., Abowd, G., Beale, R (1998). *Human-Computer Interaction*, 2nd edn, Prentice Hall, Upper Saddle River, NJ, ISBN 0-1323-9864-8.

Dray, S, Siegel, D. (2004). *Remote possibilities?: international usability testing at a distance*, Interactions, vol 11, (2), pp. 10-17.

Fidas, C., Kapsalis, V., Tranoris, C., Avouris, N., (2006). *Synchronous support and monitoring in web–based educational systems*, Journal of Campus-Wide Information Systems, vol. 23, (3), pp. 138-148, Special issue on Synchronous methods and applications in e-learning

Hartson, H , Castillo, J. , Kelso, J., Neale, W. (1996). *Remote evaluation: the network as an extension of the usability laboratory*, in Proc. of the SIGCHI conference on Human factors in computing systems: common ground, p.228-235, Vancouver, British Columbia, Canada

ISO 9241-11(1998), *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability.*

Ivory, M. Y., Hearst, M. A. (2001). *The state of the art in automating usability evaluation of user interfaces*, ACM Computing Surveys, vol 33, (4), pp. 470-516.

Krauss, F.S.H. (2003). *Methodology for remote usability activities: A case study*, IBM Systems Journal, vol 42, (4)

Nielsen J. (1993). *Usability Engineering*, Academic Press, Boston, London, ISBN 0-12-518405-0

Paganelli, L., Paterno, F. (2003). *Tools for Remote Usability Evaluation of Web Applications through Browser Logs and Task Models Behavior Research Methods*, Instruments, and Computers, vol 35, (3), pp.369-378.

Vredenburg, K., Mao, J., Smith, P., Carey, L.(2002). A *survey of user-centered design practice*, in Proc. of the SIGCHI conference on Human factors in computing systems, pp 471 -– 478, ACM Press, ISBN:1-58113-453-3

West, R, Lehman, R (2006). *Automated Summative Usability Studies: An Empirical Evaluation*, in Proc. of the SIGCHI conference on Human factors in computing systems, pp 631 – 63