Information and Software Technology 53 (2011) 789–817

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Usability evaluation methods for the web: A systematic mapping study ☆

Adrian Fernandez *, Emilio Insfran, Silvia Abrahão

ISSI Research Group, Department of Information Systems and Computation, Universitat Politécnica de Valéncia, Camino de Vera, s/n, 46022 Valencia, Spain

ARTICLE INFO

Article history:
Received 6 October 2010
Received in revised form 4 February 2011
Accepted 15 February 2011
Available online 17 March 2011

Keywords:
Usability evaluation methods
Web development
Systematic mapping

ABSTRACT

Context: In recent years, many usability evaluation methods (UEMs) have been employed to evaluate Web applications. However, many of these applications still do not meet most customers' usability expectations and many companies have folded as a result of not considering Web usability issues. No studies currently exist with regard to either the use of usability evaluation methods for the Web or the benefits they bring.

Objective: The objective of this paper is to summarize the current knowledge that is available as regards the usability evaluation methods (UEMs) that have been employed to evaluate Web applications over the last 14 years.

Method: A systematic mapping study was performed to assess the UEMs that have been used by researchers to evaluate Web applications and their relation to the Web development process. Systematic mapping studies are useful for categorizing and summarizing the existing information concerning a research question in an unbiased manner.

Results: The results show that around 39% of the papers reviewed reported the use of evaluation methods that had been specifically crafted for the Web. The results also show that the type of method most widely used was that of User Testing. The results identify several research gaps, such as the fact that around 90% of the studies applied evaluations during the implementation phase of the Web application development, which is the most costly phase in which to perform changes. A list of the UEMs that were found is also provided in order to guide novice usability practitioners.

Conclusions: From an initial set of 2703 papers, a total of 206 research papers were selected for the mapping study. The results obtained allowed us to reach conclusions concerning the state-of-the-art of UEMs for evaluating Web applications. This allowed us to identify several research gaps, which subsequently provided us with a framework in which new research activities can be more appropriately positioned, and from which useful information for novice usability practitioners can be extracted.

1. Introduction

Usability is considered to be one of the most important quality factors for Web applications, along with others such as reliability and security (Offutt [27]). Web applications are currently the backbone of business and information exchange, and are therefore the initial means to present products and services to potential customers. They are also employed by governments to disseminate relevant information to citizens. The ease or difficulty that users experience with these Web applications determines their success or failure. usability evaluation methods (UEMs) which are specifically crafted for the Web, and technologies that support the usability design process, have therefore become critical (Neuwirth and Regli [25]).

The challenge of developing more usable Web applications has led to the emergence of a variety of methods, techniques, and tools with which to address Web usability issues. Although much wisdom exists on how to develop usable Web applications, many of these applications still do not meet most customers' usability expectations (Offutt [27]). In addition, many companies have folded as a result of not considering Web usability issues (Becker and Mottay [4]). This issue has been addressed in several studies aimed at studying and/or comparing UEMs for Web development (e.g., Alva et al. [2], Cunliffe [8]). This kind of study often compares a limited number of evaluation methods, and the selection of methods is normally driven by the researcher's expectations. There is thus a need for a more systematic identification of those UEMs which have been successfully applied to Web development.

A systematic mapping study provides an objective procedure for identifying the nature and extent of the research that is available to answer a particular research question. These kinds of studies also help to identify gaps in current research in order to suggest areas for further investigation. They therefore also provide

☆ This research work is funded by the MULTIPLE project (TIN2009-13838) and the FPU program (AP2007-03731) from the Spanish Ministry of Science and Education.

* Corresponding author. Tel.: +34 96 387 73 50x83525; fax: +34 96 387 73 59.

E-mail addresses: afernandez@dsic.upv.es (A. Fernandez), einsfran@dsic.upv.es (E. Insfran), sabrahao@dsic.upv.es (S. Abrahão).

0950-5849/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.infsof.2011.02.007

a framework and background in which to appropriately develop future research activities (Budgen et al. [6]).

In previous work we have presented a preliminary systematic mapping study that was carried out to assess which UEMs have been used for Web usability evaluation and their relation to the Web development process (Insfran and Fernandez [16]). In this paper, we present an extended, improved and updated systematic mapping study with the aim of examining the current use of UEMs in Web development from the point of view of the following research question: *"What usability evaluation methods have been employed by researchers to evaluate Web artifacts, and how have these methods been used?"*. We have improved our preliminary systematic mapping study by: defining a new search string that allows more papers to be retrieved; searching in more bibliographic sources; applying new data extraction criteria and applying other synthesis techniques in order to present useful information to both researchers and practitioners.

The systematic mapping study has allowed us to outline the issues that are especially relevant to practitioners who conduct usability studies, which are, among others: how the usability evaluation methods are applied in the Web domain, what types of UEMs are most widely used, and which phase of the Web development process they are applied in. We also outline the issue that is most relevant to usability researchers, which is how to improve the current practices of Web usability research.

This paper is organized as follows. Section 2 presents the background of UEMs and introduces readers to the topic of usability evaluation. Section 3 describes the research method that was used to map the UEMs employed in Web development. Section 4 presents the results obtained from the systematic mapping study. Section 5 discusses the main findings and the limitations of this mapping study, along with the implications for research and practice. Finally, Section 6 presents our conclusions and suggests topics for further research.

2. Background

We first provide a brief background to usability evaluation methods, presenting some core ideas and several works related to UEMs. Finally, we justify the need for a systematic mapping study.

2.1. Usability evaluation methods

The term *usability* has several definitions in each research field. In the field of Human–Computer Interaction (HCI) field, the most widely accepted definition of usability is that proposed in the ISO 9241-11 [18]: *"the extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency and satisfaction in a specified context of use"*. This definition is that which is closest to the human interaction perspective. In this view, usability implies the interaction of users with the software product and can be seen as the product's capability to meet customer expectations. On the other hand, in the field of Software Engineering (SE), the most widely accepted definition of usability is that proposed in the ISO 9126-1 [20]: *"the capability of the software product to be understood, learned, operated, attractive to the user, and compliant to standards/guidelines, when used under specific conditions"*. In this view, usability is seen as one specific characteristic that affects the quality of a software product. It can be evaluated during the early stages of Web development and does not necessarily imply the user's interaction with the system since it can be measured as "conformance to specification", where usability is defined as a matter of products whose measurable characteristics satisfy a fixed specification which has been defined beforehand. These different definitions of usability directly affect how it is evaluated, since each method or technique employed in these evaluations may focus on different aspects of the term usability (e.g., effectiveness of user task, learnability of user interfaces).

A usability evaluation method is a procedure which is composed of a set of well-defined activities for collecting usage data related to end-user interaction with a software product and/or how the specific properties of this software product contribute to achieving a certain degree of usability. UEMs were formerly developed to specifically evaluate WIMP (Window, Icon, Menu, Pointing device) interfaces, which are the most representative of desktop applications. One of the most representative examples is the heuristic evaluation method proposed by Nielsen [26]. Since Web-based interfaces have grown in importance, new and adapted UEMs have emerged to address this type of user interfaces.

Although several taxonomies for classifying UEMs have been proposed, UEMs can in general terms be principally classified into two different types: *empirical methods* and *inspection methods*. Empirical methods are based on capturing and analyzing usage data from real end-users. Real end-users employ the software product (or a prototype) to complete a predefined set of tasks while the tester (human or specific software) records the outcomes of their work. Analysis of these outcomes can provide useful information to detect usability problems during the user's task completion. Inspection methods are performed by expert evaluators or designers (i.e., they do not require the participation of real end-users) and are based on reviewing the usability aspects of Web artifacts, which are commonly user interfaces, with regard to their conformance with a set of guidelines. These guidelines can range from checking the level of achievement of specific usability attributes to heuristic evaluations concerning predictions of problems related to user interfaces.

In the Web domain, both empirical and inspection methods have several advantages and disadvantages. Since the majority of Web applications are developed for many different end-user profiles, empirical methods can take into account a wide range of end-users. However, the use of empirical methods may not be cost-effective since they require a large amount of resources. Empirical methods also need a full or partial implementation of the Web application, signifying that usability evaluations are mainly moved to the last stages of the Web development process. Inspection methods, on the other hand, allow usability evaluations to be performed on Web artifacts such as mock-ups, paper prototypes, or user interface models. This is relevant because these Web artifacts can be created during the early stages of the Web development process. Another benefit of the inspection methods is that they often require fewer resources than empirical methods. However, the usability evaluation performed may be limited by the quality of the guidelines or evaluator expectations. Moreover, the interaction of real end-users is not taken into account in inspection methods.

2.2. Related work

In recent years, several studies have reported evaluations and comparisons with regard to UEMs (e.g., Gray and Salzman [12], Hartson et al. [14], Somervell and McCrickard [29]). Gray and Salzman [12] made an in-depth analysis of five experiments that compare usability evaluation methods. The aim of their study was to demonstrate that there is a definite need for scientific rigor in experiments of this type. The authors claim that most experiments on comparisons of UEMs do not clearly identify which aspects of UEMs are being compared. We agree with Gray and Salzman's criticisms, and have concluded that the results may be misleading when attempting to determine whether one UEM is more effective than another under certain conditions. However, although the studies analyzed by Gray and Salzman may be relevant in the HCI field, we consider that there is still no well-defined research method that justifies their selection of studies.

Hartson et al. [14] argue that UEMs cannot be evaluated or reliably compared since there is an important shortage of standard criteria for comparison. Several studies were analyzed in order to determine which measures had been used in the evaluation of UEMs. The majority of these studies used the thoroughness measure (the ratio between the number of real usability problems found and the number of total real usability problems). This paper showed that the majority of the comparison studies in the HCI literature on UEM effectiveness did not provide the descriptive statistics needed to perform a meta-analysis.

Somervell and McCrickard [29] presented a technique with which to evaluate heuristic evaluation methods. This study argues that the applicability of a set of heuristics to that problem set can be determined more accurately by providing the evaluators with a set of real problems. New procedures were developed in order to properly select this set of problems. Although these procedures can be applied to improve the basis of comparison for UEMs, this approach only covers a small subset of inspection methods and their applicability to specific user interfaces is *ad-hoc*.

The criticism identified in the aforementioned studies may also be applicable to the specific domain of Web interfaces. The studies that we present below are specific to the Web domain (Cunliffe [8], Ivory and Hearst [22], Alva et al. [2], Batra and Bishu [3]).

Cunliffe [8] presented an informal Web development model for mapping several UEMs proposed in literature with the stages of the Web development process. The author recognizes that the survey was not exhaustive but suggests that it could be considered as a guideline for Web designers and developers. The survey distinguishes five types of evaluation methods: competitive analysis, scenarios, inspection methods, log analysis, and questionnaires. However, several of the proposed methods are informal guidelines or means to gather information about user interaction.

Ivory and Hearst [22] published one of the most extensive studies in the field of usability evaluation. The study analyzed a large number of UEMs, taking into account their automation capability; it also proposed a taxonomy with which to classify them. UEMs are classified according to five dimensions: testing, inspection, inquiry, Analytical Modeling, and simulation. The taxonomy was applied to 128 UEMs, 58 of which were found to be suitable for Web user interfaces. The results of this survey indicate that it is important to bear in mind that the automation of usability evaluation does not capture subjective information (such as user preferences and misconceptions) since this information can only be discovered by usability testing or inquiry methods. Nevertheless, the other types of methods (Analytical Modeling and simulation) might be useful in helping designers to choose among design alternatives before committing themselves to expensive development costs. Finally, the study suggests promising ways in which to expand existing methods in order to better support automated usability evaluation.

Alva et al. [2] presented an evaluation of seven methods and tools for usability evaluation in software products and artifacts for the Web. The purpose of this study was to determine the degree of similarity among the methods using the principles defined in the ISO 9241-11 standard [18]. However, this is an informal survey with no defined research questions and no search process to identify the methods that were considered.

Batra and Bishu [3] reported the results obtained with two usability evaluation studies for Web applications. The objective of the first study was to compare the efficiency and effectiveness between user testing and heuristic evaluation. The results showed that both methods address very different usability problems and are equally efficient and effective for Web usability evaluation. The objective of the second study was to compare the performance between remote and traditional usability testing. The results indicate that there is no significant difference between the two methods.

The majority of the published studies are informal literature surveys or comparisons with no defined research questions, no search process, no defined data extraction or data analysis process, and the reviewed UEMs are selected by author criteria. In addition, the majority of these kinds of studies deal with usability evaluations in generic interfaces from any kind of system, but few studies are specifically focused on evaluation methods that have been applied to the Web domain.

2.3. Need for a systematic mapping study

Although several studies concerning UEMs have been reported, we are not aware of any systematic mapping study that has been published in the field of Web usability apart from our preliminary systematic mapping study (Insfran and Fernandez [16]). The objective of this first work was to investigate which usability evaluation methods have been used in the Web domain. A total of 51 research papers were reviewed from an initial set of 410 papers. This study provided a preliminary discussion about the methods that can be applied to evaluate the usability of Web applications.

We are aware of three studies that have been conducted in related fields (Hornbæk [15], Mendes [24], Freire et al. [11]) whose research methods belong to the evidence-based paradigm (i.e., systematic mapping studies and systematic reviews).

Hornbæk [15] applied a research method that is close to a systematic review whose aim was to review the state-of-the-practice in usability measures. The quality of the measures selected to perform usability studies was analyzed in order to investigate whether they actually measure and cover usability issues in a broad manner. This review identified several challenges in usability research such as distinguishing and empirically comparing subjective and objective usability measures, the need for developing and employing learning and retention measures, and studying correlations between usability measures as a means for validation.

Mendes [24] presented a systematic review to determine the rigor of claims of Web engineering research, demonstrating that only 5% should be considered as rigorous. The review also found that numerous papers used incorrect terminology. For instance, they used the term *experiment* rather than *experience report* or the term *case study* rather than *proof of concept*. Suggestions were proposed to improve practices in the Web Engineering field.

Freire et al. [11] presented a systematic review on Web accessibility to identify existing techniques for developing accessible content in Web applications. This review includes 53 studies, and it also proposes a classification of these techniques according to the processes described in the ISO/IEC 12207 standard [19]. This study also identified several research gaps such as considering accessibility in the use of techniques to generate Web applications based on models.

Previous work demonstrates that there is a need for a more systematic identification of which usability evaluation methods have been successfully applied to Web development. We have improved, extended and updated the systematic mapping study performed in Insfran and Fernandez [16] by: defining a new search string that allows more papers to be retrieved; searching in more bibliographic sources; applying new data extraction criteria and applying other synthesis techniques in order to present useful information to both researchers and practitioners.

3. Research method

We have performed a systematic mapping study by considering the guidelines that are provided in works as those of Budgen et al.

[6], Petersen et al. [28], and Kitchenham [23]. A systematic mapping study is a means of categorizing and summarizing the existing information about a research question in an unbiased manner.

Our systematic mapping study was performed in three stages: Planning, Conducting, and Reporting. The activities concerning the planning and conducting stages of our systematic mapping study are described in the following sub-sections and the reporting stage is presented in Section 4.

3.1. Planning stage

In this stage, we performed the following activities in order to establish a review protocol: (1) establishment of the *research question*; (2) definition of the *search strategy*, (3) *selection of primary studies*, (4) *quality assessment*, (5) definition of the *data extraction strategy* and (6) selection of *synthesis methods*. Each of them is explained in detail as follows.

3.1.1. Research question

The goal of our study is to examine the current use of UEMs in Web development from the point of view of the following research question: *"What usability evaluation methods have been employed by researchers to evaluate Web artifacts, and how have these methods been used?"*. This will allow us to categorize and summarize the current knowledge concerning Web usability evaluation, to identify gaps in current research in order to suggest areas for further investigation and to provide useful knowledge for novice usability practitioners. Since our research question is too broad, it has been decomposed into more detailed sub-questions in order for it to be addressed. Table 1 shows these research sub-questions along with their motivation.

3.1.2. Search strategy

The main digital libraries that were used to search for primary studies were: IEEEXplore, ACM Digital Library, Springer Link, and Science Direct. We also manually searched the conference proceedings and journals in which studies relevant to the Web Usability domain had previously been published:

- Conferences and workshops:
  - World Wide Web conference – WWW (2003–2009), Usability and accessibility & Web engineering tracks.
  - International conference on Web Engineering – ICWE (2003–2009).
  - International Web Usability and Accessibility workshop – IWWUA (2007–2009).
- Journals and books:
  - Internet Research Journal: "Electronic Networking Applications and Policy" - IR. Volumes 4–19 (1994–2009) (ed. Emerald Group Publishing Limited).
  - Journal of Usability Studies – JUS. Volumes 1–5 (2005–2009).
- Special issues:
  - International Journal of Human–Computer Studies "Web Usability" Special Issue – 1 volume published in 1997 (IJHCS).
  - IEEE Internet Computing Special issue on "Usability and the Web" – 1 volume published in 2002 (IEEEIC).

In order to perform the automatic search of the selected digital libraries, we used a search string (see Table 2) consisting of three parts with the aim of covering the concepts that represent the Web usability evaluation domain. The first part is related to the studies that are developed in the Web domain, the second part is related to the studies that are related to the usability domain, and the third part is related to studies that present evaluations. Table 2 shows the search string in which Boolean OR has been used to join alternate terms and synonyms in each main part; and Boolean AND has been used to join the three main parts.

The search was conducted by applying the search string to the same metadata (i.e., title, abstract and keywords) of each article for all the sources (the search string syntax was adapted in order for it to be applied in each digital library). These search terms were also taken into account in the other sources that were manually inspected in order to perform a consistent search.

The period reviewed included studies published from 1996 to 2009. This starting date was selected because 1996 was the year in which the term "Web Engineering" was coined and it has been used as starting date in other related evidence-based works in the Web domain such as that of Mendes et al. [24]. As the search was performed in 2010, publications pertaining to that year are not considered in the systematic mapping study.

In order to validate our search strategy, we compared the results obtained with a small sample of 12 primary studies (Alva et al. [S06], Atterer and Schmidt [S11], Batra and Bishu [S18],

Table 2
Search string applied.

Concept	Alternative terms & Synonyms	
Web	(web OR website OR internet OR www)	AND
Usability	(usability OR usable)	AND
Evaluation	(evalu* OR assess* OR measur* OR experiment* OR stud* OR test* OR method* OR techni* OR approach*)	

* The asterisk symbol '*' signifies any character whose purpose it is to include any word variation of each search term (e.g., the search term 'evalu*' includes the following words: evaluation OR evaluate OR evaluates OR ...).

Table 1
Research sub-questions.

Research sub-questions	Motivation
Q1. Origin of the UEMs employed	To discover whether the UEMs have been specifically crafted for the Web domain or whether they have been taken from existing UEMs from the Human–Computer Interaction field
Q2. Underlying usability definition of the UEMs employed	To discover the homogeneity in the definitions of the usability term on which the UEMs are based
Q3. Types of UEMs employed	To discover which are the most frequently employed types of UEMs, and what type of UEMs can be applied in conjunction with others
Q4. Type of evaluation performed by the UEMs employed	To discover the degree of automation that UEMs present and which usability aspects are commonly evaluated in both manual and automated evaluations
Q5. Phase(s) and Web artifacts in which the UEMs are applied	To discover during which stages of the Web development process UEMs are most frequently applied, what kind of Web artifacts that are generated during the Web development process are evaluated, and how the UEMs are integrated in the Web development processes
Q6. Feedback provided by the UEMs	To discover whether the UEMs provide recommendations and guidance to Web developers in order to overcome usability problems or whether they only provide a list of usability problems
Q7. Empirical Validation of the UEMs	To discover whether the UEMs that are proposed in the existing literature have been validated through empirical studies

Blackmon et al. [S23], Chi [S45], Conte et al. [S53], Cunliffe [S61], Hornbæk and Frøkjær [S91], Ivory and Hearst [S97], Matera et al. [S125], Molina and Toval [S130], and Olsina et al. [S142]) which we had previously identified as studies that should appear in the results in order to ensure that the search string was able to find the sample. In addition, the starting date of the search was validated by checking the references of the most relevant primary studies in order to detect whether any papers were missing. Since this validation was applied after the primary studies had been selected, this is explained in the following section.

3.1.3. Selection of primary studies

Each study that was retrieved from the automated search or the manual search was evaluated by the three authors in order to decide whether or not it should be included by considering its title, abstract and keywords. Discrepancies in the selection were solved by consensus among the three authors after scanning the entire paper. The studies that met at least one of the following inclusion criteria were included:

– Studies presenting the definition of UEM(s) that are applied to the Web domain.
– Studies reporting usability evaluations in the Web domain through the employment of existing UEM(s)

The studies that met at least one of the following exclusion criteria were excluded:

– Papers that are not focused on the Web domain.
– Papers presenting only recommendations, guidelines, or principles for Web design.
– Papers presenting only usability attributes and their associated metrics.
– Papers presenting only accessibility studies.
– Papers presenting techniques on how to aggregate usability measures.
– Papers presenting testing processes that are focused on checking functional aspects.
– Introductory papers for special issues, books and workshops.
– Duplicate reports of the same study in different sources.
– Papers not written in English.

The references of the selected studies (only those which had been found to be most relevant by each digital library) were followed in order to check whether other relevant studies could be included in our search. This procedure allowed us to validate the starting date of our systematic mapping study. Although relevant studies related to the usability evaluation domain were found (e.g. Nielsen [26]), no relevant studies specifically focused on the Web domain were found prior to 1996.

The reliability of inclusion of a candidate study in the systematic mapping study was assessed by applying Fleiss' Kappa [10]. Fleiss' Kappa is a statistical measure for assessing the reliability of agreement between a fixed number of raters when classifying items. This measure is scored as a number between 0 (poor agreement) and 1(full agreement). We asked three independent raters to classify a random sample of 20 studies, 10 of which had previously been included in the mapping study and 10 of which had not. The Fleiss' kappa obtained was 0.84. This indicates an acceptable level of agreement among raters.

3.1.4. Quality assessment

A three-point Likert-scale questionnaire was designed to provide a quality assessment of the selected studies. The questionnaire contained three subjective closed-questions and two objective closed-questions. The subjective questions were:

(a) The study presents a detailed description of the UEM employed.
(b) The study provides guidelines on how the UEM can be applied.
(c) The study presents clear results obtained after the application of the UEM.
The possible answers to these questions were: "I agree (+1)", "Partially (0)", and "I do not agree (−1)".
The objective questions were as follows:
(d) The study has been published in a relevant journal or conference proceedings. The possible answers to this question were: "Very relevant" (+1), "Relevant (0)", and "Not so relevant (−1)". This question was rated by considering the order of relevance provided by the digital library, the CORE conference ranking (A, B and C conferences), and the Journal Citation Reports (JCR) lists.
(e) The study has been cited by other authors. The possible answers to this question were: "Yes (+1)" if the paper has been cited by more than five authors; "Partially (0)" if the paper has been cited by 1–5 authors; and "No (−1)" if the paper has not been cited. This question was rated by considering the Google scholar citations count. It is important to note that the minimum score for early publications (i.e., papers published in 2009) is considered as "Partially (0)" in order not to penalize them.

Each of the studies selected has a score for each closed-question that has been calculated as the arithmetic mean of all the individual scores from each reviewer. The sum of the five closed-question scores of each study provides a final score (an integer between −5 and 5). These scores were not used to exclude papers from the systematic mapping study but were rather used to detect representative studies in order to discuss each research sub-question.

3.1.5. Data extraction strategy

The data extraction strategy that was employed was based on providing the set of possible answers for each research sub-question that had been defined. This strategy ensures the application of the same extraction data criteria to all selected papers and it facilitates their classification. The possible answers to each research sub-question are explained in more detail as follows.

With regard to Q1 (*Origin of the UEMs employed*), a paper can be classified in one of the following answers:

(a) *New*: if it presents at least one evaluation method that is specifically crafted for the Web.
(b) *Existing*: if the paper uses existing methods from the HCI field in the Web domain.

With regard to Q2 (*Underlying usability definition of UEMs employed*), a paper can be classified in one of the following answers:

(a) *Standard*: if the underlying usability definition of the UEM is based on standards such as 9241-11 [18] or ISO 9126-1 [20].
(b) *Ad-hoc*: if the underlying usability definition of the UEM is based on an *ad-hoc* definition by other authors.

With regard to Q3 (*Types of UEMs employed*), the taxonomy proposed by Ivory and Hearst [22] was employed in order to classify the UEMs. A paper can be classified in one or more of the following answers:

(a) *Testing*: if it involves an evaluator observing participants interacting with a user interface to determine usability problems (e.g., think-aloud protocol, remote testing, log file analysis).

(b) *Inspection*: if it involves an expert evaluator using a set of criteria to identify potential usability problems (e.g., heuristic evaluation, guideline reviews, or cognitive walkthroughs).
(c) *Inquiry*: if it presents a method that gathers subjective input from participants, such as their preferences or their feelings (e.g., focus group, interviews, and questionnaires).
(d) *Analytical Modeling*: if it presents an engineering approach that enables evaluators to predict usability by employing different kinds of models (e.g., GOMS analysis, Cognitive Task Analysis).
(e) *Simulation*: if it simulates user interaction through any kind of simulation algorithm or the analysis of usage data (e.g. Petri net models, information scent).

With regard to Q4 (*Type of evaluation performed by the UEMs*), a paper can be classified in one of the following answers:

(a) *Automated*: if it presents a tool that automatically performs the entire method or a large portion of the method (e.g., log analyzers, source code or model checkers, user simulators). This means that the evaluator only needs to interpret the results since the main evaluation tasks are performed automatically.
(b) *Manual*: if it presents a usability evaluation that is performed manually, signifying that the method can be computer-aided but that the main evaluation tasks need to be performed by a human evaluator (e.g., interviews, user questionnaires, think-aloud methods).

With regard to Q5 (*Phase(s) and Web artifacts in which the UEMs are applied*), a paper can be classified in one or more ISO/IEC 12207 [19] high-level processes:

(a) *Requirements*: if the artifacts that are used as input for the evaluation include high-level specifications of the Web application (e.g., task models, uses cases, usage scenarios).
(b) *Design*: if the evaluation is conducted on the intermediate artifacts that are created during the Web development process (e.g., navigational models, abstract user interface models, dialog models).
(c) *Implementation*: if the evaluation is conducted at the final user interface or once the Web application is completed.

With regard to Q6 (*Feedback provided by the UEMs*), a paper can be classified in one of the following answers:

(a) *Yes*: if the UEM provides recommendations or guidance to the designer on how the detected usability problems can be corrected.
(b) *No*: if the UEM is aimed at only reporting usability problems.

With regard to Q7 (*Empirical Validation of the UEMs*), a paper can be classified in one of the following types of strategies that can be carried out depending on the purpose of the validation and the conditions for empirical investigation (Fenton and Pfleeger [9]):

(a) *Survey*: if it provides an investigation performed in retrospect, when the method has been in use for a certain period of time in order to obtain feedback about the benefits and limitations of the UEM.
(b) *Case study*: if it provides an observational study in which data is collected to evaluate the performance of the UEM throughout the study.
(c) *Controlled experiment*: if it provides a formal, rigorous, and controlled investigation that is based on verifying hypotheses concerning the performance of the UEM.
(d) *No*: if it does not provide any type of validation or if it only presents a proof of concept.

In order to validate our data extraction strategy, the Fleiss' Kappa statistic [10] was applied to assess the agreement among evaluators when the studies were classified into the possible answers. We asked three independent raters to classify a random sample of 15 studies that had previously been included in the review. Average Fleiss' kappas for each research sub-question were: Q1: 0.84; Q2: 0.95; Q3: 0.79; Q4: 0.93; Q5: 0.81; Q6: 0.83 and Q7: 0.81. Overall, this result suggests an acceptable level of agreement among raters.

A template for both quality assessment and data extraction activities was designed to make easier the management of the data extracted for each paper (see Appendix B).

3.1.6. Synthesis methods

We applied both quantitative and qualitative synthesis methods. The quantitative synthesis was based on:

- Counting the primary studies that are classified in each answer from our research sub-questions.
- Defining bubble plots in order to report the frequencies of combining the results from different research sub-questions. A bubble plot is basically two x–y scatter plots with bubbles in category intersections. This synthesis method is useful to provide a map and giving a quick overview of a research field (Petersen et al. [28]).
- Counting the number of papers found in each bibliographic source per year.

The qualitative synthesis is based on:

- Including several representative studies for each research sub-question by considering the results from the quality assessment.
- Summarizing the benefits and limitations of the UEMs classified in each proposed research sub-question.

3.2. Conducting stage

The application of the review protocol yielded the following preliminary results (see Table 3):

A total of 206 research papers were therefore selected in accordance with the inclusion criteria. We found several issues at this stage:

- Some studies had been published in more than one journal/conference. In this case, we selected only the most complete version of the study.
- Some studies appeared in more than one source. In this case, they were taken into account only once according to our search order, which was the following: IEEEXplore, ACM, Springer Link, Science Direct, etc.

The search results revealed that the research papers concerning Web usability had been published in several conferences/journals related to different fields such as Human–Computer Interaction (HCI), Web Engineering (WE), and other related fields.

4. Results

The overall results, which are based on counting the primary studies that are classified in each of the answers to our research sub-questions, are presented in Table 4. Any readers who wish to

Table 3
Results of the conducting stage.

Source	Potential studies	Selected Studies
Automated search		
IEEExplore (IEEE)	863	83
ACM DL (ACM)	960	63
Springer Link (SL)	571	16
Science Direct (SD)	179	11
Total	2573	173
Manual search		
WWW Conference	46	5
ICWE Conference	32	7
IWWUA Workshop	20	4
Internet Research Journal	11	4
Journal of Usability Studies	9	5
International Journal of HCS	7	1
IEEE Internet Computing	5	3
Other	-	4
Total	130	33
Overall results from both searches	2703	206

Table 4
Results of the systematic mapping.

Research sub-questions	Possible answers	Results	
		# Studies	Percentage (%)
Q1. Origin of the UEMs employed	New	81	39.32
	Existing	125	60.68
Q2. Underlying usability definition of the UEMs employed	Standard	37	17.96
	Ad-hoc	169	82.04
Q3. Types of UEMs employed	User testing	121	58.74
	Inspection	88	42.72
	Inquiry	72	34.95
	Analytical Modeling	44	21.36
	Simulation	17	8.25
Q4. Type of evaluation performed by the UEMs employed	Manual	143	69.42
	Automated	63	30.58
Q5. Phase(s) and Web artifacts in which the UEMs are applied	Requirements	7	3.40
	Design	53	25.73
	Implementation	187	90.78
Q6. Feedback provided by the UEMs	Yes	65	31.55
	No	141	68.45
Q7. Empirical Validation of the UEMs	Survey	25	12.14
	Case Study	32	15.53
	Experiment	34	16.50
	No	115	55.83

Note that Q3 and Q5 are not exclusive; a study can be classified in one or more of the answers. The summation of the percentages is therefore over 100%.

view the complete list of selected studies included in this systematic mapping study are referred to Appendix A. Both the classification of the selected papers in each category and their quality scores are provided in Appendix C.

The following sub-sections present the analysis of the results from each research sub-question, the map created by combining different sub-questions, and to what extent the UEMs for the Web domain may be an interest topic after analyzing the number of research studies for each year covered.

4.1. Origin of the UEMs employed

The results for sub-question Q1 (*Origin of the UEMs employed*) revealed that around 39% of the papers reviewed had usability evaluation methods that were specifically designed for the Web (see Table 4). For instance, we found representative examples of these methods in Blackmon et al. [S23], Conte et al. [S53], and Triacca et al. [S185].

Blackmon et al. [S23] proposed the Cognitive Walkthrough for the Web method (CWW). CWW is an adaptation of the original Cognitive Walkthrough (CW) method. Since CWW was crafted for applications that support use by exploration, CWW is presented as an appropriate method for the evaluation of Web sites. The aim of CWW is to simulate users performing navigation tasks on a Web site by assuming that the users perform goal-driven exploration.

Conte et al. [S53] presented the Web Design Perspectives method (WDP). This method extends and adapts the generic heuristics for user interfaces proposed by Nielsen [26] with the aim of drawing closer to the dimensions that characterize a Web application: content, structure, navigation and presentation.

Triacca et al. [S185] proposed a usability inspection method for Web applications called the Milano–Lugano Evaluation Method (MiLE+). This method distinguishes between the application-independent analysis and the application-dependent analysis. The former is related to a technical and objective perspective, whereas the latter is related to the specific context of use of the Web application and how it meets user goals.

The remaining 61% of the studies reported the use of existing evaluation methods from the HCI field such as cognitive walkthroughs, heuristic evaluations, questionnaires or remote user testing (see Table 4). These methods have been defined to be applied in any kind of user interfaces without considering the application domain. These results may indicate that there are more UEMs adapted from existing methods to be applied in the Web domain than UEMs that have been defined by considering the specific characteristics of Web applications. We observed that the UEMs for the Web pay special attention to content and navigational issues, and not only to the user behavior. This fact is relevant since the main dimensions that define Web applications are content, navigation and presentation. We consider that UEMs for the Web should address the usability concept in a broader manner by considering usability aspects that are related to the aforementioned dimensions, and not only focus on usability aspects related to the effectiveness and efficiency of users in performing tasks, or the end-user satisfaction.

4.2. Underlying usability definition of the UEMs

The results for sub-question Q2 (*Underlying usability definition of the UEMs*) revealed that around 82% of the papers reviewed present UEMs that are based on an *ad-hoc* definition of the usability concept (see Table 4). On the other hand, around 18% of the papers reviewed present UEMs whose definition of the usability concept is based on standards (see Table 4). For instance, we found representative examples of these methods in Alonso-Rios et al. [S04], Moraga et al. [S131], and Oztekin et al. [S144].

Alonso-Rios et al. [S04] presented an HTML analyzer that parses HTML code in order to extract usability information from Web pages. This analyzer basically examines usability aspects which are related to ease of navigation, understandability, flexibility and compatibility, and these are based on the World Wide Web Consortium (W3C) guidelines [30]. These aspects are classified into six categories related to the Web application source code (i.e., Web page, images, forms, tables, lists and links).

Moraga et al. [S131] presented a UEM for evaluating second generation Web portals (i.e., portlets). This method is based on a usability model that decomposes usability into measurable concepts and attributes. The measurable concepts (e.g., understandability, learnability) of this usability model are based on the usability sub-characteristics proposed in the quality model of the ISO 9126-11 standard [20].

Oztekin et al. [S144] proposed the UWIS methodology for usability assessment and design of Web-based information systems. UWIS is a checklist whose aim is to provide usability indexes. These usability indexes are defined by considering the usability sub-characteristics proposed in the ISO 9241-11 [18] (i.e., effectiveness, efficiency and satisfaction), the dialogue principles for user interface design according to the ISO 9241-10 [17] standard, and the usability heuristics proposed by Nielsen [26].

The results for this sub-question indicate that the UEMs are based on different underlying concepts of usability. This raises several issues, since these UEMs may not evaluate the same aspects of usability. The comparison of UEMs in order to determine their performance is therefore considered to be a complex task. This problem results from the fact that the usability concept has not been homogeneously defined. Although several approaches present UEMs whose usability definition is based on standards, these standards are not consistent with each other. This could be alleviated, at least to some extent, if new proposals consider the next generation of standards (i.e., ISO 25000 SQuaRE standard in progress [21]) in order to define the aspects of usability to be evaluated. The SQuaRE standard integrates both perspectives of the usability concept: *usability of the software product* which is based on the ISO 9126-1 standard; and *usability in use* which is based on the ISO 9241-11 standard. This provides a comprehensive structure for the role of usability as part of software quality (Bevan [5]).

4.3. Types of UEMs employed

The results for sub-question Q3 (*Types of UEMs employed*) revealed that the most frequently used type of UEM is *user testing*, signifying that around 59% of the papers reviewed reported some kind of testing involving users (see Table 4). These results may indicate that most evaluations are performed during the later stages of the Web development lifecycle. We identified the following representative sub-types of user testing methods:

– *Think-Aloud Protocol*: users think aloud while they are performing a set of specified tasks. Examples of this UEM sub-type are reported in works such as Krahmer and Ummelen [S118], Stefano et al. [S171], and Van Waes [S188].
– *Question-Asking Protocol*: testers ask the users direct questions. Examples of this UEM sub-type are reported in the studies conducted by Corry et al. [S56], Gee [S75] and Wang and Liu [S193].
– *Performance Measurement*: testers or software tools record usage data and obtain statistics during the test. Examples of this UEM sub-type are reported in works such as Nakamichi et al. [S134], Nakamichi et al. [S135] and Norman and Panizzi [S138].
– *Log Analysis*: testers or software tools analyze usage data. Examples of this UEM sub-type are reported in works such as Chi [S45], Costagliola and Fuccella [S58], and Kazienko and Pilarczyk [S110]. When usage data is particularly related to gaze points obtained from the analysis of eye movement, the method is called *Eye Tracking*. Examples of Eye Tracking methods are reported in works such as Cooke and Cuddihy [S55], and De Kock et al. [S63].
– *Remote Testing*: Testers and users are not co-located during the test. These methods are commonly applied in conjunction with *Log Analysis* methods. Examples of this UEM sub-type are reported in works such as Lister [S121], Paganelli and Paterno [S146], and Thompson et al. [S180].

Inspection methods account for around 43% of the papers reviewed (see Table 4). Although inspection methods are intended to be performed by expert evaluators, most of them were applied by novice evaluators such as Web designers or students in order to compare the results. We identified the following representative sub-types of inspection methods:

– *Heuristic evaluation:* experts identify heuristic violations in Web artifacts. Examples of this UEM sub-type are reported in works such as Allen et al. [S03], Nielsen and Loranger [S136], and Oztekin et al. [S144].
– *Cognitive Walkthrough:* experts simulate a user's goal achievement by going through a set of tasks. Examples of this UEM sub-type are reported in works such as Clayton et al. [S52], and Filgueiras et al. [S69]. Core ideas of cognitive walkthroughs have led to the emergence of concrete methods for the Web domain such as the Cognitive Walkthrough for the Web (Blackmon et al. [S23]), and the Metaphor of Human-Thinking (Hornbæk and Frøkjær [S91]).
– *Perspective-based inspection*: experts conduct an oriented and narrow evaluation that can be based on design perspectives, inspectors' tasks, or metric calculation. Some examples of this sub-type of methods are the Web Design Perspectives (Conte et al. [S53]), the Abstract-Tasks Inspection (Costabile and Matera [S57]), and the WebTango Methodology (Ivory and Hearst [S98]).
– *Guideline review*: experts verify the consistency of Web artifacts by using a set of usability guidelines. Examples of this UEM sub-type are reported in works such as Becker and Mottay [S20] and Vanderdonckt et al. [S189].

Inquiry methods account for around 35% of the papers reviewed (see Table 4). Since these methods focused on gathering subjective data from users, the majority were used in combination with other types of methods such as testing or inspection to perform a more complete evaluation. We identified the following representative sub-types of inquiry methods:

– *Questionnaire:* users provide answers to specific questions. Examples of this UEM sub-type are reported in works such as Cao et al. [S37] and Zaharias [S202].
– *Interviews:* One user and one expert participate in a discussion session concerning the user's attitude towards the artifact to be evaluated. Examples of this UEM sub-type are reported in works such as Van Velsen et al. [S187] and Vatrapu and Pérez-Quiñones [S190].
– Focus group: Multiple users participate in a discussion session concerning their attitudes towards the artifact to be evaluated. Examples of this UEM sub-type are reported in works such as Go et al. [S77] and Jung et al. [S105].

Analytical Modeling accounts for around 21% of the papers reviewed (see Table 4). This is intended to model certain aspects such as user interfaces, task environments, or user performance in order to predict usability. We identified the following representative sub-types of Analytical Modeling methods:

– *Cognitive Task Analysis*: User tasks are modeled in order to predict usability problems. Examples of this UEM sub-type are reported in works such as Paganelli and Paterno[S145], and Saward et al. [S158].
– *Task environment analysis*: Evaluation of the mapping between users' goals and user interface tasks. Examples of this UEM sub-type are reported in works such as Ahn et al. [S02] and Bolchini et al. [S29].
– *GOMS analysis*: Human task performance is modeled in terms of Goals, Operators, Methods, and Selection rules (GOMS) in order to predict execution and learning time. Examples of this UEM sub-type are reported in works such as Tonn-Eichstädt [S184].

Simulation methods only account for around 8% of the papers reviewed (see Table 4). Few methods can be considered to be only simulation methods, since they present characteristics from other kinds of methods (particularly from Analytical Modeling). These are mainly based on agents or algorithms whose intention is to simulate user behavior. For example, Chi et al. [S46] presented the Information Scent Absorption Rate, which measures the navigability of a Website by computing the probability of users reaching their desired destinations on the Web site. The InfoScent Bloodhound Simulator tool was developed to support this method with the aim of generating automated usability reports. This paper presents a user study which argues that Bloodhound correlates with real users surfing for information on four Websites and that it can reduce the need for human work during usability testing.

4.4. Type of evaluation performed by the UEMs

The results for sub-question Q4 (*Type of evaluation performed by the UEMs*) revealed that around 69% of the studies performed the evaluations manually, whereas around 31% of the studies reported the existence of some kind of automated tool to support the proposed method (see Table 4). These tools are mainly based on source code checking, usage data or log analysis, and user simulation. Some examples of automated evaluations were found in Becker and Berkemeyer [S19], Ivory and Megraw [S99], and Vanderdonckt et al. [S189]

Becker and Berkemeyer[S19] proposed a technique to support the development of usable Web applications. This technique is supported by a GUI-based toolset called Rapid Application Design and Testing (RAD–T) which allows early usability testing during the design stage. Usability evaluations are possible since Self-Testing Hypertext Markup Language (ST-HTML) was developed as an HTML extension in order to integrate usability and functional requirements into Web page items. These requirements can be verified through an inspection of the ST-HTML source code.

Ivory and Megraw [S99] proposed the WebTango methodology. The purpose was to define a set of quantitative measures and compute them for a large sample of rated Web interfaces. Data obtained from these computations can be used to derive statistical models from the measures and ratings. This approach not only allows the statistical models to be employed to predict ratings for new Web interfaces, but the significance of the measures can also be evaluated. A tool was developed to automate various steps of this methodology, such as obtaining of the statistical models or the calculation of certain measures.

Vanderdonckt et al. [S189] proposed a usability evaluation method based on the automated review of guidelines. Usability and accessibility guidelines from literature were interpreted and expressed in the Guideline Definition Language (an XML-compliant formal language). In this approach, a guideline can be evaluable if HTML elements reflect its semantics. These guidelines mainly focus on aspects such as color combinations, alternative text for visual content, etc. A tool was developed to illustrate how these formal guidelines can be checked in Web page source code.

The results for this sub-question indicate that the majority of the efforts in automated UEMs are focused on the source code since it is the only artifact employed in most cases. There is a shortage of this kind of methods which can evaluate, for example, intermediate artifacts such as abstract user interfaces or navigational models. Most of the tools found are based on the operationalization of usability guidelines (mostly focused on aesthetic issues), or on calculating and interpreting usability measures at the final user interface level. However, it is important to note that automated usability evaluation has several drawbacks. It is oriented towards gathering objective data, hence, user perceptions and user context, cannot be considered. Although automated UEMs can reduce efforts and resources, they should be used in conjunction with other UEMs in order to consider as many usability dimensions as possible.

4.5. Phase(s) and Web artifacts in which the UEMs are applied

The results for sub-question Q5 *(Phases and Web artifacts in which the UEMs are applied*) revealed that around 90% of the evaluations are performed at the implementation level of the Web application (see Table 4). This kind of usability evaluations is also known as summative evaluation. It takes place after the product has been developed, or possibly when a prototype version is ready. The artifacts that were most commonly analyzed were the final Web user interfaces and the logs that contain the user actions. For instance, Nakamichi et al. [S135] presented the WebTracer tool for recording and analyzing the user's operations on Web pages while they directly interact with the website. The aim was to collect quantitative data to detect possible usability problems without interrupting the user's operation.

Around 26% of the studies (see Table 4) describe evaluations performed at the design level, employing the intermediate artifacts obtained during the Web development process (e.g., abstract user interfaces, navigational models). This kind of usability evaluations is also known as formative evaluation. For instance, Atterer and Schmidt [S11] proposed a prototype of a model-based usability validator. The aim was to perform an analysis of models that represent enriched user interfaces. This approach takes advantage of navigational and presentation models that are available in model-driven Web development methods (e.g., WebML [7] or OO–H [13]) since they contain data concerning the ways in which the site is intended to be traversed and abstract properties of the page layout.

Only around 3% of the studies (see Table 4) describe evaluations performed at the *requirements* specification level (e.g., laboratory user testing of paper mock-ups or prototypes). One representative example was found in Molina and Toval [S130] who suggested integrating usability requirements in the development of model-driven Web applications is presented. The aim is to extend the expressiveness of the models that define the navigation of the Web application in order to represent usability requirements that can be evaluated through the application of automated metrics.

The results for this sub-question indicate that there is a need for UEMs that can be used at early stages of the Web development lifecycle. Although evaluations at the implementation stage are necessary to explore user behavior, since there are usability aspects that can only be accessed through user interaction, applying UEMs only at this stage can lead to various difficulties since more of them may be detected later. Correcting these problems can make the maintenance of the source code difficult. Usability evaluations must be performed not only at the implementation stage but also during each phase of the Web application development. If usability problems are detected earlier, the quality of the final Web applications can be improved, thus saving resources in the implementation stage. This could contribute towards a reduction in the cost of the Web development process.

4.6. Feedback provided by the UEMs

The results for sub-question Q6 (*feedback provided by the UEMs*) revealed that around 68% of the studies only provided reports on usability problems, giving no explicit feedback and guidance to the corresponding design activities. The remaining studies (around 32%) also offered suggestions for design changes based on the usability problems detected (see Table 4). Some representative examples of this were found in Blackmon et al. [S24], Chi [S45] and Hornbæk and Frøkjær [S92].

Blackmon et al. [S24] reported two experiments aimed at presenting Cognitive Walkthrough for the Web (CWW) as an effective UEM with which to repair usability problems related to unfamiliar and confusable links. CWW uses the Latent Semantic Analysis algorithm (LSA) to compute the semantic similarities between the user goals and the headings/links/descriptions of other widgets. This enables developers to very quickly check whether the Web application links are also comprehensible and not confusing for their intended users, and if not, it provides guidance on how to repair them.

Chi[S45] presented a visualization method based on data mining for Web applications. The purpose is to apply a set of techniques in order to help developers to understand usage data, content changes and linkage structures. These techniques can be used to identify specific usability problems on large Web sites, where they discover major traffic patterns and propose changes to improve how the user accesses the Web content. The ScentViz prototype was developed to implement these techniques and to show how usability evaluations can be enhanced using visualization methods.

Hornbæk and Frøkjær [S92] reported on an experiment aimed at comparing the assessment of both the usability and utility of problems, and redesign suggestions. The results of the experiment showed how redesign proposals were assessed by developers as being of higher utility than simple problem descriptions. Usability problems were seen more as a help in prioritizing ongoing design decisions.

The results for this sub-question indicate that most of the UEMs have been designed to generate a list of usability problems, but not to provide explicit guidance on how these problems can be properly corrected. Usability evaluation must take into account both activities: discovering and repairing usability problems. Simply employing lists of usability problems is not sufficient. The developers need more support to explore new alternatives with which to improve their designs. This indicates a need for new UEMs or extensions of existing methods to incorporate redesign issues as an integral part of the evaluation method. If this goal is to be attained, the evaluation methods need to be integrated into the Web development process to a greater extent in order to understand the traceability between the usability problems detected and the artifacts that originate these usability problems.

4.7. Empirical Validation of the UEMs

The results for sub-question Q7 (*Empirical Validation of the UEMs*) revealed that 56% of the studies did not conduct any type of validation of the method (see Table 4). Around 12% of the studies presented UEMs which had been validated through a survey (see Table 4). For instance, Zaharias [S202] proposed a questionnaire for evaluating e-learning applications. Two pilot trials were conducted and analyzed in order to validate the coverage of the questionnaire. Results obtained from the empirical evaluation allowed new versions of the questionnaire to be developed in order for it to be more reliable.

Around 16% of the papers report case studies (see Table 4). For instance, Matera et al. [S125] presented a case study in which three methods were applied to the evaluation of a Web application: design inspections to examine the hypertext specification, Web usage analysis to analyze user behavior, and a heuristic evaluation to analyze the released prototypes and the final Web application. The case study took place in an iterative development process, in which versions of Web applications were released, evaluated, and improved by taking into account the problems encountered during the evaluation.

Around 17% of the papers report controlled experiments (see Table 4). For instance, Bolchini and Garzotto [S30] performed an empirical study to evaluate the quality of the MiLE+ method. The concept of quality was operationalized into attributes in order to facilitate the measuring process. These attributes were: the degree to which the method supports the detection of all usability problems (performance) and how fast this detection (efficiency) takes place; the effort needed by an evaluator to perform an evaluation with the method (cost-effectiveness) and the ease with which the method was learnt (learnability). The results show that the MiLE+ method achieved acceptable levels in all attributes, providing a good support for inexperienced evaluators. However, this experiment was conducted solely with experts and novice users, and the results obtained were not compared with other methods, making it difficult to draw conclusions as to why this method should be used rather than others.

The results for this sub-question show that experiments (17%) and case studies (16%) were the most frequently employed types of empirical methods used for validation purposes. This is explained by the fact that experimentation is a common research method in the Human–Computer Interaction field, and case studies are commonly used in the Software Engineering field. However, since only 44% of the papers included validations, there would appear to be a need for more validation studies.

4.8. Mapping results

The seven research sub-questions were combined in order to establish a mapping with the aim of providing an overview of the Web usability evaluation field. This mapping allows us to obtain more information about how the results from each sub-question are related to the others, and what the possible research gaps are.

Fig. 1a shows the mapping results obtained from research sub-questions Q1 (*Origin*) and Q2 (*Usability definition*) in comparison to research sub-questions Q5 (*Stages*) and Q7 (*Validation*). These results may indicate that:

- The majority of UEMs that are specifically crafted for the Web are applied at the implementation stage of the Web development process and present more empirical validations than the UEMs that were taken from the HCI field.
- The majority of UEMs whose underlying usability definition is based on standards are likely to present more empirical validations compared with the number of UEMs whose underlying usability definition is based on ad-hoc definitions. However, the majority of these UEMs are applied in later stages of the Web development process.

Fig. 1b shows the mapping results obtained from research sub-questions Q1 (*Origin*) and Q2 (*Usability definition*) in comparison to research sub-questions Q4 (*Type of evaluation*) and Q6 (*Feedback*). These results may indicate that:

- Fewer UEMs adapted from existing HCI methods have been automated than UEMs developed specifically for the Web
- Most UEMs have been designed to report only a list of usability problems, independent of their origin or underlying usability definition.

Fig. 1c shows the mapping results obtained from research sub-questions Q5 (*Stages*) and Q7 (*Validation*) in comparison to research sub-questions Q4 (*Type of evaluation*) and Q3 (*Type of UEM*). These results may indicate that:

– The majority of automated UEMs are applied at the implementation stage, where the most common method is user testing. However, inspection methods are likely to be used at earlier stages of the Web development process, especially in the design stage.
– There is a need to perform more empirical validations of the UEMs, regardless of the type of method and the type of evaluation performed.

Fig. 2a shows the mapping results obtained from research sub-question Q3 (*type of UEM*) when compared with itself. These results may indicate that:

– UEMs are not used in isolation since it is a common practice to apply several different UEMs in order to address a broad range of usability problems.
– Inquiry methods are likely to be combined with user testing and inspection methods in order to provide subjective feedback from users.

Fig. 2b shows the mapping results obtained from research sub-questions Q1 (*Origin*), Q2 (*Usability definition*) and Q3 (*stages*) when combined. These results may indicate that:

– There is a shortage of UEMs whose usability definition is based on standards, regardless of their origin or type of method.
– The majority of UEMs that are specifically crafted for the Web are defined as inspection, user testing and Analytical Modeling methods.

Fig. 2c shows the mapping results obtained from research sub-questions Q3 (*Type of UEM*), Q4 (*Type of evaluation*) and Q6 (*Feedback*) when combined. These results may indicate that:

– User testing methods are likely to be more automated than the other types of usability evaluation methods.
– Only few automated methods provide explicit recommendations and guidance to Web developers in comparison to the manual usability evaluation methods.

4.9. Interest of the topic

Web usability evaluation has led to the appearance of a large number of studies in recent years. These studies can be found in papers published mainly in the fields of Human–Computer Interaction and Web Engineering. All the studies agree on the importance of usability evaluations in the Web domain. However, the scope of most of the studies found is centered on reporting the usability evaluation results of a specific Web application. There are fewer studies with a broad scope, implying that almost none of the pa-

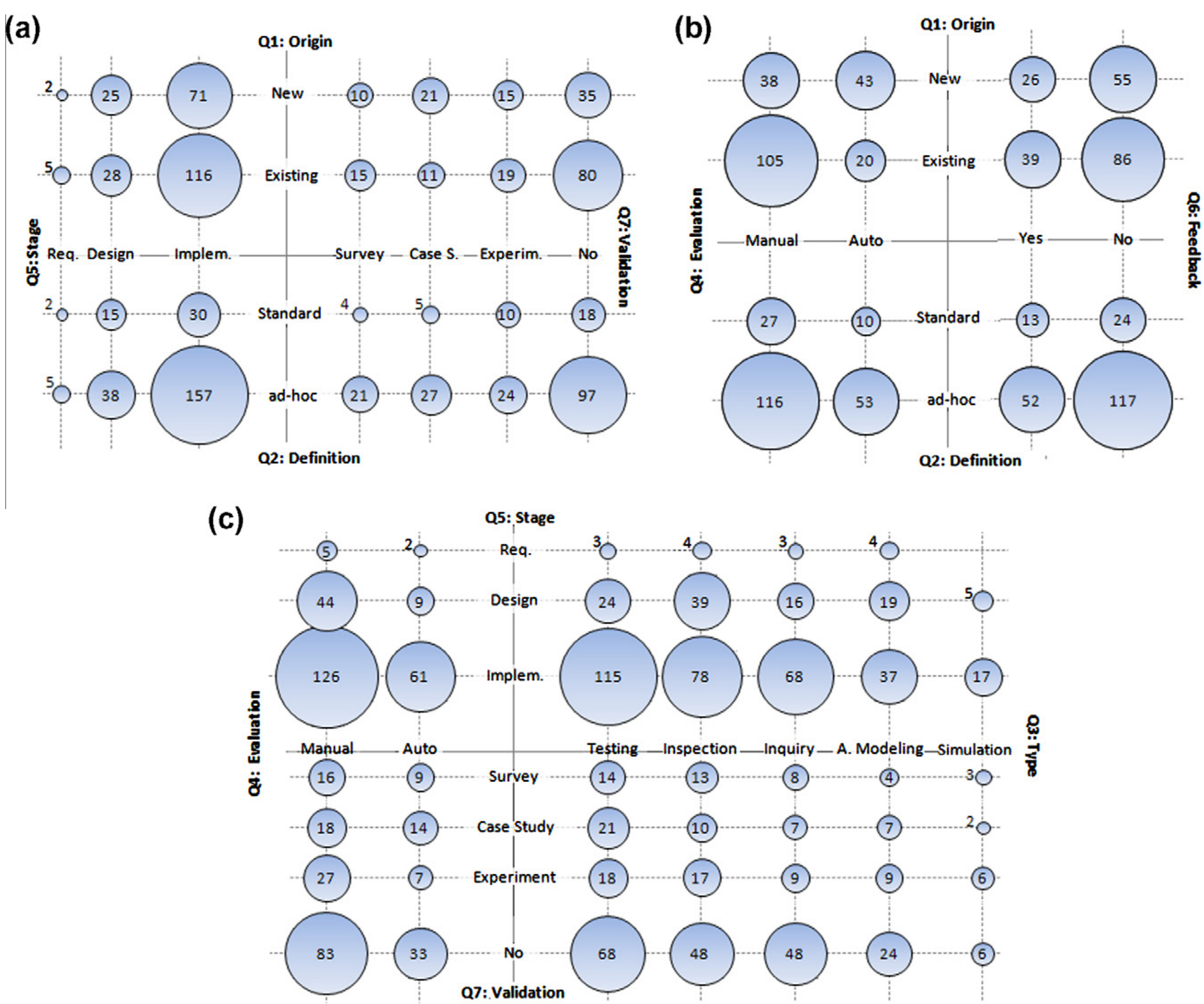

Fig. 1. Mapping results obtained from the combination of research sub-questions (I).

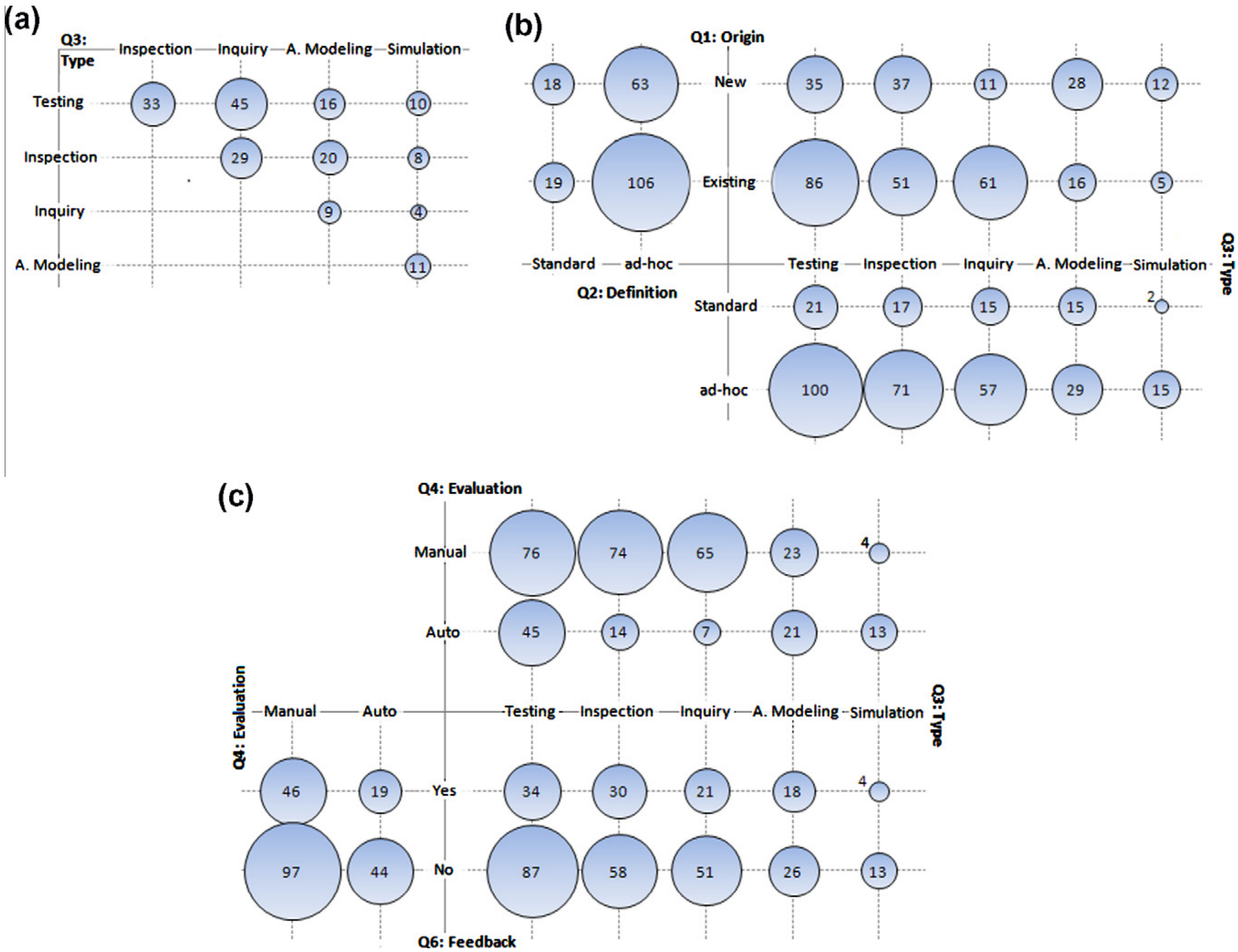

Fig. 2. Mapping results obtained from the combination of research sub-questions (II).

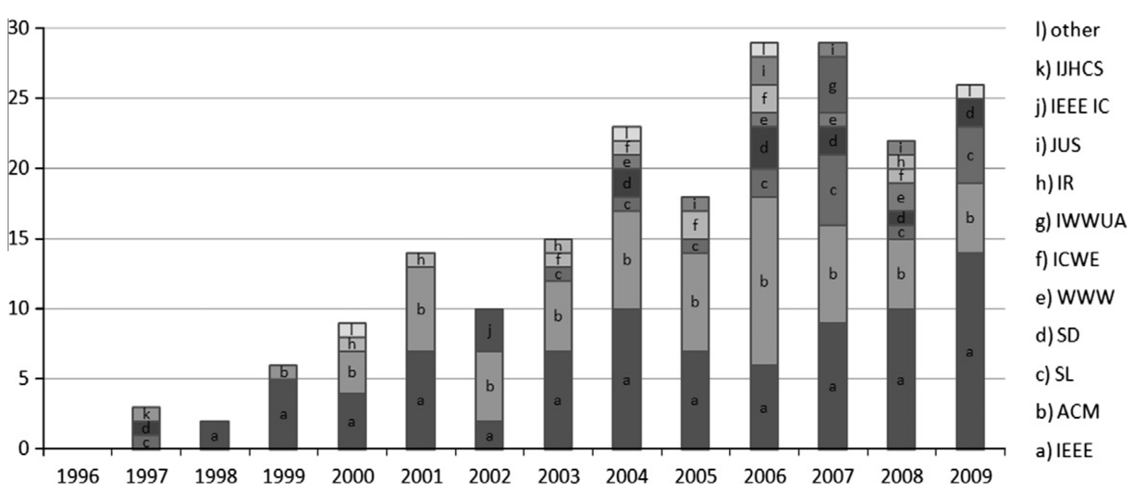

Fig. 3. Number of publications on Web usability by year and source.

pers provided results that can be generalized for a particular Web vertical domain (e.g., e-commerce, e-government, e-learning).

Fig. 3 shows the number of selected publications on Web usability evaluation methods by year and source. The analysis of the number of research studies on Web usability showed that there has been a growth of interest in this topic, particularly since 2004.

The relative increase in this topic was of about 766% (from three selected studies in 1997 up to 26 selected studies in 2009). This can be considered as an indicator of how usability evaluation methods for the Web have gained importance in recent years. The following terms: *Software Engineering*, *Web Engineering*, *Human–Computer Interaction*, and *Usability Evaluation* were also sought in the same

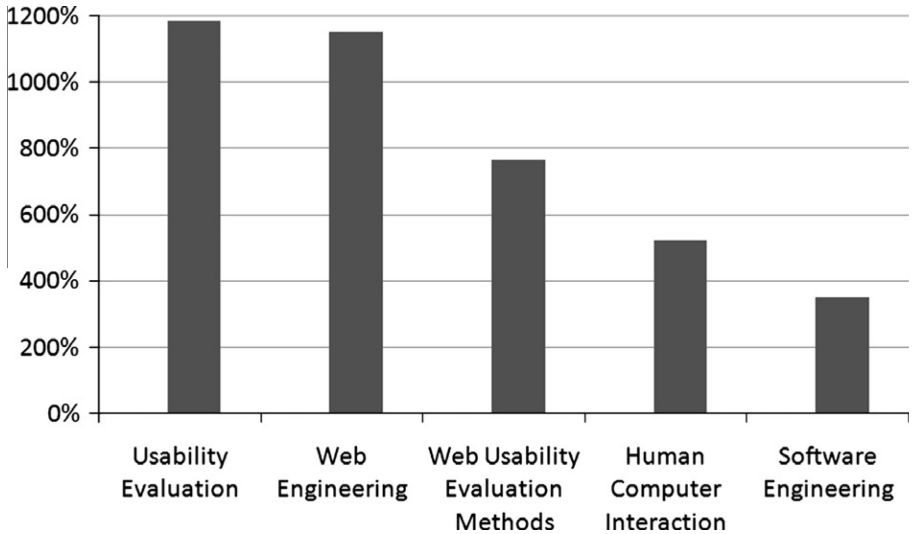

Fig. 4. Relative increase means associated to related research fields.

digital libraries that were selected in our search strategy with the objective of obtaining the relative increase mean associated with these research fields. Fig. 4 shows a comparison of these relative increases with that obtained from our systematic mapping study. Since the Web usability evaluation method topic can be considered as a sub-topic of *Usability evaluation* and *Web engineering*, these results confirm the interest in the topic.

There are no conclusions with regard to which the best bibliographic sources are since those papers that appeared in several sources were considered only once. However, most of the relevant studies concerning usability evaluation methods applied to Web domain were found in the IEEExplore and the ACM digital library.

5. Discussion

This section summarizes the principal findings of this systematic mapping study. It also highlights the limitations that may represent threats to its validity and discusses the implications for research and practice.

5.1. Principal findings

The goal of this systematic mapping study was to examine the current use of usability evaluation methods in Web development. The principal findings of our study are the following:

- Usability evaluation methods have been constantly modified to better support the evaluation of Web artifacts. However, the methods evaluate different usability aspects depending on the underlying definition of the usability concept (ISO 9241-11 [18], ISO 9126-1 [20]). Therefore, there is no single method that is suitable for all circumstances and type of Web artifacts. It depends on the purpose of the evaluation and the type of artifact that is evaluated (e.g., abstract user interfaces, log files, final Web user interfaces). Our results suggest that a combination of methods (e.g., inspection and inquiry methods) could provide better results.
- The majority of the papers reported on evaluations at the implementation phase (e.g., final user interfaces, log analysis). The study also reveals that the evaluations are mainly performed in a single phase of the Web application development.
- There is a shortage of automated evaluation methods, specifically those that can be applied at early stages (e.g. requirements specifications, navigational models, presentation models).
- The majority of the papers do not present any kind of validation. Among the papers that present empirical validations, several controlled experiments have been reported. More replications are therefore needed to build up a body of knowledge concerning usability evaluation methods for the Web.
- The majority of the methods reviewed only reported a list of usability problems; they did not provide explicit feedback or suggestions to help designers improve their artifacts.
- Web usability evaluation is an important topic and interest in it is growing.

5.2. Limitations of the systematic mapping study

The principal limitations of this systematic mapping study are related to publication bias, selection bias, inaccuracy in data extraction, and misclassification. Publication bias refers to the problem that positive results are more likely to be published than negative ones since negative results take longer to be published or are cited in other publications to a lesser extent (Kitchenham [23]). In order to alleviate this threat (at least to some extent), we scanned relevant special issues of journals and conference proceedings. However, we did not consider grey literature (i.e., industrial reports or PhD theses) or unpublished results. This may have affected the validity of our results to some extent.

Selection bias refers to the distortion of a statistical analysis owing to the criteria used to select publications. We attempted to alleviate this threat (at least to some extent) by defining our inclusion criteria in order to gather the largest possible amount of papers that fit into the Web usability evaluation domain.

Inaccuracy in data extraction and misclassification refer to the possibility of a study's information being extracted in different ways by different reviewers. In order to alleviate this threat (at least to some extent), the extraction and classification of the papers was conducted by all three authors. Each of the 206 studies was classified by each reviewer and the discrepancies that appeared were solved by consensus.

We have also detected other limitations related to the systematic mapping procedure itself. Since the goal of systematic mapping studies is more oriented towards categorizing the selected papers and identifying representative studies rather than performing evidence aggregations of empirical results, the results of empirical validations should be analyzed by considering more specific research questions (e.g., *how many unique usability evaluation methods have not been validated?*, *which usability evaluation methods have proven* *to be the most effective?*). This could be done by applying aggregation techniques in order to combine evidence, although these techniques are more commonly applied in systematic reviews.

5.3. Implications for research and practice

The findings of our systematic mapping study have implications for both researchers who are planning new studies of usability evaluations of Web applications and for practitioners who are working in Web development companies and would like to integrate usability evaluation methods into their Web development process in an effective manner.

For researchers, we believe that the usability concept has not been defined consistently in the ISO standards (as shown in Table 4, Fig. 1a and b), which might be a problem since usability as a quality characteristic may not actually cover all the usability aspects even though the UEMs used are effective. We therefore consider that new UEMs should take into account all the usability definitions and specific Web application characteristics in order to provide more complete results.

Our findings show that the majority of the papers reported evaluations at the implementation phase or in a single phase of the Web application development (as shown in Table 4 and Fig. 1a and c). Usability evaluations at each phase of the Web application development are critical to ensure that the product will actually be usable. We therefore consider that there is an important shortage of evaluation methods with which to address usability in the early stages of Web application development, and not only when the application is partially or fully implemented. The main problem seems to be that most Web development processes do not take advantage of the intermediate artifacts that are produced during early stages of the Web development process (i.e., requirements and design stages). These intermediate artifacts (e.g., navigational models, abstract user interface models, dialog models) are mainly used to guide developers and to document the Web application. Since the traceability between these artifacts and the final Web application are not well-defined, performing evaluations using these artifacts can be difficult. New research should be oriented towards integrating usability evaluations into the Web development process whose intermediate artifacts can be effectively evaluated. For instance, this problem does not appear in model-driven Web development processes in which models (intermediate artifacts) that specify an entire Web application are applied in all the steps of the development process, and the final source code is automatically generated from these models (Abrahão et al. [1]). The evaluation of these models can provide early usability evaluation reports in order to suggest changes that can be directly reflected in the source code. Our study confirms the viability of this approach, since some papers applied usability evaluations in model-driven development processes (e.g., Atterer and Schmidt [S11], Molina and Toval [S130]). This is also reflected in most of the automated evaluation methods that were found which also perform evaluations of artifacts obtained during the implementation phase such as HTML source code (as shown in Fig. 1c). Research into automated evaluation methods should go further. It should also be focused on the evaluation of intermediate artifacts applied at early stages (e.g. requirements specifications, navigational models, presentation models).

A further finding was that the majority of the reviewed methods only allowed the generation of a list of usability problems (as shown in Table 4, Figs. 1b and 2c). There is little guidance or suggestions to help designers with the problem of how the usability problems can be corrected. UEMs need to include suggestions about how the identified usability problems can be corrected.

Finally, we detected that few validations of UEMs have been published in literature (as shown in Table 4 and Fig. 1a and c). When a method is proposed, it is essential to conduct experiments to provide empirical evidence about its usefulness (e.g. ease of use, effectiveness, efficiency, application cost). More controlled experiments are therefore needed to compare the proposed methods. They should use the same measures in order to determine which methods are the most appropriate in different situations.

We have also learned some lessons that may be useful for practitioners. These lessons are related to which kind of UEM can be applied at different stages of the Web development process and how they can be combined.

Owing to the fact that few UEMs are applied at the requirements analysis stage, we could only draw conclusions about the design and implementation stages. The types of methods that were most widely applied at the design stage were Inspection methods (as shown in Fig. 1b). These methods focus mainly on evaluating abstract or partially implemented user interfaces. They are mainly based on heuristic evaluation and guideline reviews that do not require end-user participation. This makes them useful for application by Web developers themselves; however, in most cases these evaluations need to be performed by expert evaluators. The types of methods that were most frequently applied at the implementation stage were User testing, Inspection, and Inquiry methods (as shown in Fig. 1b). These methods mainly focus on evaluating the final Web application or usage data log. Both types require user participation and their planning is often more costly than heuristic evaluations.

Table 5 suggests several usability evaluation methods by considering the results obtained from the quality assessment, along with the results obtained from the answers to each research subquestion. The rows of the table show each UEM and the columns show the answers for each criterion from the extracted data. Practitioners who are interested in performing usability studies by using these UEMs can refer to the attached references.

Practitioners must bear in mind that there is no single UEM that addresses all the existing usability problems. Most of the studies therefore employed more than one UEM in order to take advantage of the evaluation infrastructure. For instance, in most cases in which a user testing method was applied (e.g., Think-Aloud Protocol, Remote Testing), it was often combined with another inquiry method (e.g., Questionnaires, Focus Group, Interviews), thereby taking full advantage of end-user participation in order to gather both objective and subjective data (see Fig. 2a).

An important task for practitioners is not only to compare results from different UEMs, but also to collect data concerning the employment of the UEMs, that can be used to assess the usability of the UEM itself. This data can be very useful in detecting deficiencies and in re-designing evaluation methods in order for them to be more effective and easier to apply.

6. Conclusions and further work

In recent years, a great number of methods have been employed to evaluate the usability of Web applications. However, no mapping studies exist that summarize the benefits and drawbacks of UEMs for the Web domain since the majority of studies are informal literature surveys driven by the researcher's expectations.

This paper has presented a systematic mapping study that summarizes the existing information regarding usability evaluation methods that have been employed by researchers to evaluate Web artifacts. From an initial set of 2703 papers, a total of 206 research papers were selected for the mapping study, and the results obtained have allowed us to extract conclusions regarding the state-of-the-art in the field, to identify several research gaps, and to extract some guidelines for novice usability practitioners. Moreover, the application of a well-defined review protocol will also allow us to efficiently update and extend the systematic mapping study in future years.

Table 5
Usability evaluation methods that may be of interest to practitioners.

Ref.	UEM	Origin	Definition	Type (s)	Automated	Stage (s)	Feedback	Empirically validated	Quality Score
Blackmon et al. [S23]	Cognitive Walkthrough for the Web	New	Ad-hoc	Inspection Analytical Simulation	Yes	Design	Yes	Experiment (Blackmon et al. [S24])	5
Chi et al. [S46]	InfoScent Simulator	New	Ad-hoc	Testing Simulation	Yes	Implem.	No	Experiment (itself)	5
Conte et al. [S53]	Web Design Perspectives	New	Standard	Inspection	No	Design	No	Experiment (itself)	5
Costabile and Matera [S57]	Abstract-Tasks (Systematic Usability Evaluation)	New	Ad-hoc	Inspection	No	Implem.	Yes	Experiment (itself)	5
Hornbæk and Frøkjær [S91]	Metaphor of Human-Thinking	Exis.	Ad-hoc	Inspection	No	Design Implem.	Yes	Experiment (Hornbæk and Frøkjær [S91][S92])	5
Ivory and Hearst [S98]	WebTANGO	New	Ad-hoc	Inspection Analytical	Yes	Design Implem.	No	Survey (Ivory and Megraw [S99])	5
Nakamichi et al. [S135]	WebTracer	New	Ad-hoc	Testing	Yes	Implem.	No	Case study (itself)	4
Nielsen and Loranger [S136]	Web Heuristic Evaluation	New	Ad-hoc	Inspection	No	Design Implem.	Yes	Survey (itself)	5
Triacca et al. [S185]	MILE+	New	Standard	Inspection	No	Design	Yes	Experiment (Bolchini and Garzotto [S30])	4
Van Waes [S188]	Think-Aloud Protocol	Exis.	Ad-hoc	Testing	No	Implem.	Yes	Experiment (Krahmer and Ummelen [S118], Van Velsen et al. [S187])	5
Zaharias [202]	Questionnaire	Exis.	Ad-hoc	Inquiry	No	Implem	No	Survey (itself and Cao et al. [S37]) Experiment (Van Velsen et al. [186])	5

The results obtained show the need for usability evaluation methods that are specifically crafted for the Web domain, which can be better integrated into the Web application lifecycle, particularly during the early stages of the Web development process.

We hope that our findings will be useful in the promotion and improvement of the current practice of Web usability research, and will provide an outline to which usability evaluation methods can be applied in order to evaluate Web artifacts and how they are employed.

Since this systematic mapping study has been performed by considering solely those research papers published in digital libraries, journals and books, our future work will include other usability studies conducted in industrial settings (e.g., reports available online). There is also a need for more in-depth analyses of the level of integration of UEMs into the different processes of the Web application lifecycle.

Ongoing research is based on performing a systematic review with the aim of addressing the following research question: *which usability evaluation methods have proven to be the most effective?* For this task, we intend to employ the knowledge obtained in this systematic mapping study in order to select the most relevant empirical studies concerning UEMs and to extract data regarding comparisons between different UEMs. Future research is also planned to define a usability evaluation method for Web applications that are designed using model-driven development. Considering the inherent advantages of model-driven development, usability problems will be detected by evaluating the intermediate artifacts (models) generated during different stages of the Web development process, particularly in the early stages.

Appendix A. Primary studies selected

[S001] Aberg, J. and Shahmehri, N. 2001. "An empirical study of human Web assistants: implications for user support in Web information systems". Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI'01), pp. 404–411. [ACM]

[S002] Ahn, J. Brusilovsky, P. He, D. Grady, J. and Li, Q. 2008. "Personalized Web Exploration with Task Models". Proc. of the 17th international conference on World Wide Web (WWW '08), pp. 1–10. [WWW]

[S003] Allen, M. Currie, L. Bakken, S. Patel, V. and Cimino, J. 2006. "Heuristic evaluation of paper-based Web pages: A simplified inspection usability methodology". Journal of Biomedical Informatics, Volume 39, Issue 4, pp. 412–423. [SD]

[S004] Alonso-Rios, D. Luis-Vazquez, I. Mosqueira-Rey, E. Moret-Bonillo, V. and Del Rio, B.B. 2009. "An HTML analyzer for the study of web usability," Proc. of the IEEE International Conference on Systems, Man and Cybernetics (SMC 2009), pp. 1224–1229. [IEEEx]

[S005] Alshamari, M. and Mayhew, P. 2008. "Task Design: Its Impact on Usability Testing". Proc. of the Third International Conference on Internet and Web Applications and Services (ICIW '08), pp. 583–589. [IEEEx]

[S006] Alva, M. Martínez, A. Cueva, J.M. Sagástegui, C. and López, B. 2003. "Comparison of Methods and Existing Tools for the Measurement of Usability in the Web". Proc. of the 3rd International Conference on Web Engineering (ICWE'03), pp. 386–389. [ICWE]

[S007] Al-Wabil, A. and Al-Khalifa, H. 2009. "A framework for integrating usability evaluations methods: The

(*continued on next page*)

Appendix A (*continued*)

Mawhiba web portal case study". Proc. of the International Conference on the Current Trends in Information Technology (CTIT'09), pp. 1–6. [IEEEx]

[S008] Anandhan, A. Dhandapani, S. Reza, H. and Namasivayam, K. 2006. "Web Usability Testing – CARE Methodology". Proc. of the Third International Conference on Information Technology: New Generations (ITNG'06), pp. 495–500. [IEEEx]

[S009] Ardito, C. Lanzilotti, R. Buono, P. and Piccinno, A. 2006. "A tool to support usability inspection". Proc. of the Working Conference on Advanced Visual Interfaces (AVI '06), pp. 278–281. [ACM]

[S010] Arroyo, E. Selker, T. and Wei, W. 2006. "Usability tool for analysis of web designs using mouse tracks". Proc. of the Conference on Human Factors in Computing Systems, pp. 484–489. [ACM]

[S011] Atterer, R. and Schmidt, A. 2005. "Adding Usability to Web Engineering Models and Tools". Proc. of the 5th International Conference on Web Engineering (ICWE'05), pp. 36–41 [ICWE]

[S012] Atterer, R. Wnuk, M. and Schmidt, A. 2006. "Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction". Proc. of the 15th international conference on World Wide Web (WWW'06), pp. 203–212. [WWW]

[S013] Atterer, R. and Schmidt, A. 2007. "Tracking the interaction of users with AJAX applications for usability testing". Proc. of the SIGCHI conference on Human factors in computing systems (CHI'07), pp. 1347–1350. [ACM]

[S014] Bachiochi, D. Berstene, M. Chouinard, E. Conlan, N. Danchak, M. Furey, T. Neligon, C. and Way, D. 1997. "Usability studies and designing navigational aids for the World Wide Web". Computer Networks and ISDN Systems, Vol. 29, Issues 8-13, pp. 1489–1496. [SD]

[S015] Badre, A. and Jacobs, A. 1999. "Usability, aesthetics, and efficiency: an evaluation in a multimedia environment". Proc. of IEEE International Conference on Multimedia Computing and Systems, Vol.1, pp. 103–106. [IEEEx]

[S016] Bartell, A.L. 2005. "Using content analysis and Web design heuristics to evaluate informational Web sites: an exploratory study". Proc. of the International Professional Communication Conference (IPCC'05), pp. 771-777. [IEEEx]

[S017] Basu, A. 2003. "Context-driven assessment of commercial Web sites". Proc. of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03), pp. 8–15. [IEEEx]

[S018] Batra, S. Bishu, R.R. 2007. "Web usability and evaluation: issues and concerns". Proc. of the 2nd international conference on Usability and internationalization (UI-HCII'07), pp. 243–249. [ACM]

[S019] Becker, S.A. and Berkemeyer, A. 2002. "Rapid application design and testing of Web usability". IEEE Multimedia, vol. 9, no. 4, pp. 38–46. [IEEEx]

[S020] Becker, S.A. and Mottay, F.E. 2001. "A global perspective on Web site usability". IEEE Software, vol. 18, no. 1, pp. 54–61. [IEEEx]

[S021] Bednarik, R. Gerdt, P. Miraftabi, R. and Tukiainen, M. 2004. "Development of the TUP model – evaluating educational software". Proc. of the IEEE International Conference on Advanced Learning Technologies (ICALT'04), pp. 699–701. [IEEEx]

[S022] Bevis, K.J. and Henke, K.A. 2008. "Evaluating usability in an information product". Proc. of the IEEE International Professional Communication Conference (IPCC'08), pp. 1–5. [IEEEx]

[S023] Blackmon, M.H. Polson, P.G. Kitajima, M. and Lewis, C. 2002. "Cognitive walkthrough for the web". Proc. of the SIGCHI conference on Human factors in computing systems (CHI'02), pp. 463–470. [ACM]

[S024] Blackmon, M.H. Kitajima, M. and Polson, P.G. 2003. "Repairing usability problems identified by the cognitive walkthrough for the web". Proc. of the SIGCHI conference on Human factors in computing systems (CHI'03), pp. 497–504. [ACM]

[S025] Blackmon, M.H. Kitajima, M. and Polson, P.G. 2005. "Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs". Proc. of the SIGCHI conference on Human factors in computing systems (CHI'05), pp. 31–40. [ACM]

[S026] Blake, C.T. and Rapanotti, L. 2004. "Usability evaluation of distributed groupware in distance learning". Proc. of the Fifth International Conference on Information Technology Based Higher Education and Training (ITHET'04), pp. 500–504. [IEEEx]

[S027] Bojko, A. 2006. "Using Eye Tracking to Compare Web Page Designs: A Case Study". Journal of Usability Studies, Vol. 1, Issue 3, pp. 112–120. [JUS]

[S028] Bok-Lee, K. and Grice, R.A. 2003. "An adaptive viewing application for the web on personal digital assistants". Proc. of the 21st annual international conference on Documentation (SIGDOC'03), pp. 125–132. [ACM]

[S029] Bolchini, D. Paolini, P. and Randazzo, G. 2003. "Adding hypermedia requirements to goal-driven analysis". Proc. of the 11th IEEE International Requirements Engineering Conference, pp. 127–137. [IEEEx]

[S030] Bolchini, D. and Garzotto, F. "Quality of Web Usability Evaluation Methods: An Empirical Study on MiLE+". Proc. of the International Workshop on Web Usability and Accessibility (IWWUA'07), pp. 481–492. [IWWUA]

[S031] Brajnik, G. 2000. "Automatic Web Usability Evaluation: Where Need to be Done?". Proc. of the 6th Conference on Human Factors and the Web. Available at: <http://users.dimi.uniud.it/~giorgio.brajnik/papers/hfweb00.html>. [Other]

[S032] Brajnik, G. Cancila, D. Nicoli, D. and Pignatelli, M. 2005. "Do text transcoders improve usability for disabled users?". Proc. of the International Cross-Disciplinary Workshop on Web Accessibility (W4A'05), pp. 9–17. [ACM]

[S033] Burmeister, O.K. 2000. "Usability testing: revisiting informed consent procedures for testing internet sites". Proc. of the second Australian Institute conference on Computer ethics, pp. 3–9. [ACM]

[S034] Burton, C. and Johnston, L. 1998. "Will World Wide Web user interfaces be usable?". Proc. of the Computer Human Interaction Conference (OZCHI'98), pp. 39–44. [IEEEx]

[S035] Burton, M.C. and Walther, J.B. 2001. "A survey of web log data and their application in use-based design". Proc. of the 34th Annual Hawaii International Conference on System Sciences, pp. 10. [IEEEx]

Appendix A (*continued*)

[S036] Cakir, C.B. and Oztaysi, B. 2009. "A model proposal for usability scoring of websites," Proc. of the International Conference on Computers & Industrial Engineering (CIE'09), pp. 1418–1422. [IEEEx]
[S037] Cao, J. Crews, J.M. Nunamaker, J.F. Jr. Burgoon, J.K. and Lin, M. 2004. "User experience with Agent99 Trainer: a usability study". Proc. of the 37th Annual Hawaii International Conference on System Sciences, pp. 11. [IEEEx]
[S038] Carstens, D.S. and Patterson, P. 2005. "Usability Study of Travel Websites". Journal of Usability Studies, Vol. 1, Issue 1, pp. 47–61. [JUS]
[S039] Chadwick-Dias, A. McNulty, M. and Tullis, T. 2003. "Web usability and age: how design changes can improve performance". Proc. of the Conference on Universal Usability (CUU '03), pp. 30–37. [ACM]
[S040] Chandrashekar, S. Stockman, T. Fels, D. and Benedyk, R. 2006. "Using think aloud protocol with blind users: a case for inclusive usability evaluation methods". Proc. of the 8th international ACM SIGACCESS conference on Computers and accessibility (Assets '06), pp. 251–252. [ACM]
[S041] Chang, W. Hon, S. and Chu, C. 2003. "A systematic framework for evaluating hyperlink validity in Web environments". Proc. of the Third International Conference on Quality Software (QSIC'03), pp. 178–185. [IEEEx]
[S042] Chatley, R. Kramer, J. Magee, J. and Uchitel, S. 2003. "Visual methods for Web application design". Proc. of the IEEE Symposium on Human Centric Computing Languages and Environments (HCC'03), pp. 242–244. [IEEEx]
[S043] Chattratichart, J. and Brodie, J. 2004. "Applying user testing data to UEM performance metrics": Proc. of the Conference on Human factors in computing systems (CHI '04), pp. 1119–1122. [ACM]
[S044] Cheng-ying, M. and Yan-sheng, L. 2004. "Testing and evaluation for Web usability based on extended Markov chain model". Wuhan University Journal of Natural Sciences, Vol. 9, No. 5, pp. 687–693. [SL]
[S045] Chi, E. 2002. "Improving Web Usability Through Visualization". IEEE Internet Computing Vol. 6, Issue 2, pp. 64–71. [IEEE IC]
[S046] Chi, E. Rosien, A. Supattanasiri, G. Williams, A. Royer, C. Chow, C. Robles, E. Dalal, B. Chen, J. and Cousins, S. 2003. "The bloodhound project: automating discovery of web usability issues using the InfoScent simulator". Proc. of the SIGCHI conference on Human factors in computing systems (CHI '03), pp. 505–512. [ACM]
[S047] Choros, K. and Muskala, M. 2009. "Block Map Technique for the Usability Evaluation of a Website". Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems, LNCS 5796, Springer, pp. 743–751. [SL]
[S048] Chou, E. 2002. "Redesigning a large and complex website: how to begin, and a method for success". Proc. of the 30th annual ACM SIGUCCS conference on User services (SIGUCCS '02), pp. 22–28. [ACM]
[S049] Chun-Long, M. and Mendes, E. 2005. "Web Usability Measurement: Comparing Logic Scoring Preference to Subjective Assessment". Proc. of the 5th International Conference on Web Engineering (ICWE'05), pp. 53–62. [ICWE]
[S050] Çınar, M.O. 2009. "Eye Tracking Method to Compare the Usability of University Web Sites: A Case Study". Human Centered Design, LNCS 5619, Springer, pp. 671–678. [SL]
[S051] Clark, J. Van Oorschot, P.C. and Adams, C. 2007. "Usability of anonymous web browsing: an examination of Tor interfaces and deployability". Proc. of the 3rd symposium on Usable privacy and security (SOUPS '07), pp. 41–51. [ACM]
[S052] Clayton, N. Biddle, R. and Tempero, E. 2000. "A study of usability of Web-based software repositories". Proc. of the International Conference on Software Methods and Tools (SMT'00), pp. 51–58. [IEEEx]
[S053] Conte, T. Massollar, J. Mendes, E. and Travassos, G.H. 2009. "Web usability inspection technique based on design perspectives". IET Software, Vol.3, No.2, pp. 106–123. [IEEEx]
[S054] Cooke, L. 2004. "Improving usability through eye tracking research". Proc. of the International Professional Communication Conference (IPCC'04), pp. 195–198. [IEEEx]
[S055] Cooke, L. and Cuddihy, E. 2005. "Using eye tracking to address limitations in Think-Aloud Protocol". Proc. of the International Professional Communication Conference (IPCC'05), pp. 653–658. [IEEEx]
[S056] Corry, M.D. Frick, T.W. and Hansen, L. 1997. "User-centered design and usability testing of a web site: An illustrative case study". Educational Technology Research and Development, Vol. 45, No. 4, pp. 65–76. [SL]
[S057] Costabile, M.F. and Matera, M. 2001. "Guidelines for hypermedia usability inspection". IEEE Multimedia, Vol. 8, No. 1, pp. 66–69. [IEEEx]
[S058] Costagliola, G. and Fuccella, V. 2009. "A visual system for analyzing user behaviour in web tasks". Proc. of the IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'09), pp. 101–104. [IEEEx]
[S059] Cuddihy, E. Wei, C. Barrick, J. Maust, B. Bartell, A.L. and Spyridakis, J.H. 2005. "Methods for assessing web design through the internet". Proc. of the Human factors in computing systems (CHI '05), pp. 1316–1319. [ACM]
[S060] Cugini, J. and Scholtz, J. 1999. "VISVIP: 3D visualization of paths through web sites". Proc. of the International Workshop on Database and Expert Systems Applications, pp. 259–263. [IEEEx]
[S061] Cunliffe,D. 2000. "Developing Usable Web Sites – A Review and Model". Internet Research Journal, Vol. 10, Issue 4, pp. 295–307. [IR]
[S062] De Angeli, A. Sutcliffe, A. and Hartmann, J. 2006. "Interaction, usability and aesthetics: what influences users' preferences?". Proc. of the 6th Conference on Designing Interactive systems (DIS '06), pp. 271–280. [ACM]
[S063] De Kock, E. Van Biljon, J. and Pretorius, M. 2009. "Usability evaluation methods: mind the gaps". Proc. of

(*continued on next page*)

Appendix A (*continued*)

the Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT '09), pp. 122–131. [ACM]
[S064] De Marsico, M. and Levialdi, S. 2004. "Evaluating web sites: exploiting user's expectations". International Journal of Human-Computer Studies, Vol. 60, Issue 3, pp. 381–416. [SD]
[S065] De Wet, L. Blignaut, P. and Burger, A. 2002. "Comprehension and usability variances among multicultural web users in South Africa". Proc. of the SIGCHI Conference on Human factors in computing systems (CHI '02), pp. 526–527. [ACM]
[S066] Douglas, I. 2006. "Collaborative International Usability Testing: Moving from Document-based Reporting to Information Object Sharing". Proc. of the International Conference on Global Software Engineering (ICGSE '06), pp.114–118. [IEEEx]
[S067] Duan, J. and Zhang, N. 2007. "Research on Visualization Techniques for Web Usability Analysis". Proc. of the Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'07), Vol. 2, pp. 788–791. [IEEEx]
[S068] El Helou, S. Gillet, D. Salzmann, C. and Yu, C.M. 2009. "A Study of the Acceptability of a Web 2.0 Application by Higher-Education Students Undertaking Collaborative Laboratory Activities". Proc. of the Second International Conference on Advances in Computer–Human Interactions (ACHI '09), pp. 117–125. [IEEEx]
[S069] Filgueiras, L. Martins, S. Tambascia, C. and Duarte, R. 2009. "Recoverability Walkthrough: An Alternative to Evaluate Digital Inclusion Interfaces," Proc. of the Latin American Web Congress (LE-WEB'09), pp. 71–76. [IEEEx]
[S070] Fisher, J. and Burstein, F. 2008. "Usability + usefulness = trust: an exploratory study of Australian health web sites". Internet Research Journal, Vol. 18, No. 5, pp. 477–498. [IR]
[S071] Fraternali, P. and Tisi, M. 2008. "Identifying Cultural Markers for Web Application Design Targeted to a Multi-Cultural Audience". Proc. of the 8th International Conference on Web Engineering (ICWE'08), pp. 231–239. [ICWE]
[S072] Fuhrmann, S. Bosley, J. Li, B. Crawford, S. MacEachren, A. Downs, R. and Gahegan, M. 2003. "Assessing the usefulness and usability of online learning activities: MapStats for kids". Proc. of the annual national conference on Digital government research (dg.o '03), pp. 1–4. [ACM]
[S073] García, E. Sicilia, M.A. González, L.A. and Hilera, J.R. 2003. "A Concept-Based Approach for the Design of Web Usability Evaluation Questionnaires". Web Engineering, LNCS 2722, Springer, pp. 407–410. [SL]
[S074] García, F. Lozano, M. Montero, F. Gallud, J. González, P. and Lorenzo, C. 2006. "A Controlled Experiment for Measuring the Usability of WebApps Using Patterns". Enterprise Information Systems VII, Part 5, pp. 257–264. [SL]
[S075] Gee, K. 2001. "The ergonomics of hypertext narative: usability testing as a tool for evaluation and redesign". Journal of Computer Documentation (JCD) , Vol. 25, Issue 1, pp. 3–16. [ACM]

Appendix A (*continued*)

[S076] Georgiakakis, P. Retalis, S. Psaromiligkos, Y. and Papadimitriou, G. 2007. "Depth Toolkit: A Web-Based Tool for Designing and Executing Usability Evaluations of E-Sites Based on Design Patterns". Human-Computer Interaction. Interaction Design and Usability, LNCS 4550, Springer, pp. 453–462. [SL]
[S077] Go, K. Takahashi, T. and Imamiya, A. 2000. "A case study on participatory redesign of web site with scenario-based techniques". Proc. of the Seventh International Conference on Parallel and Distributed Systems: Workshops (PADSW'00), pp. 161–166. [IEEEx]
[S078] González, M.P. Lorés, J. and Granollers, A. 2007. "Assessing Usability Problems in Latin-American Academic Webpages with Cognitive Walkthroughs and Datamining Techniques". Usability and Internationalization. HCI and Culture, LNCS 4559, Springer, pp. 306–316. [SL]
[S079] Grady, H.M. 2000. "Web site design: a case study in usability testing using paper prototypes". Proc. of Joint IEEE International Professional Communication Conference and 18th Annual Conference on Computer Documentation (IPCC/SIGDOC 2000), pp. 39–45. [IEEEx]
[S080] Granic, A. and Glavinic, V. 2006. "Evaluation of interaction design in web-based intelligent tutoring systems". Proc. of the 28th International Conference on Information Technology Interfaces, pp. 265–270. [IEEEx]
[S081] Granic, A. Mitrovic, I. and Marangunic, N. 2008. "Usability evaluation of web portals". Proc. of the 30th International Conference on Information Technology Interfaces (ITI'08), pp. 427–432. [IEEEx]
[S082] Habuchi, Y. Kitajima, M. and Takeuchi, H. 2008. "Comparison of eye movements in searching for easy-to-find and hard-to-find information in a hierarchically organized information structure". Proc. of the symposium on Eye tracking research & applications (ETRA '08), pp. 131–134. [ACM]
[S083] Han, M. and Park, P. 2009. "A study of interface design for widgets in web services through usability evaluation". Proc. of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human (ICIS '09), pp. 1013–1018. [ACM]
[S084] Harper, S. Yesilada, Y. Goble, C. and Stevens, R. 2004. "How much is too much in a hypertext link?: investigating context and preview – a formative evaluation". Proc. of the fifteenth ACM conference on Hypertext and hypermedia (HYPERTEXT '04), pp. 116–125. [ACM]
[S085] Harrison, T.M. Zappen, J.P. and Watson, D. 2009. "Children's use of government information systems: design and usability". Proc. of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government (dg.o '09), pp. 113–122. [ACM]
[S086] Hart, J. Ridley, C. Taher, F. Sas, C. and Dix, A. 2008. "Exploring the facebook experience: a new approach to usability". Proc. of the 5th Nordic conference on Human-computer interaction: building bridges (NordiCHI '08), pp. 471–474. [ACM]

Appendix A (*continued*)

[S087] Hart, D. and Portwood, D.M. 2009. "Usability testing of web sites designed for communities of practice: tests of the IEEE Professional Communication Society (PCS) web site combining specialized heuristic evaluation and task-based user testing". Proc of the IEEE International Professional Communication Conference (IPCC'09), pp. 1–17. [IEEEx]

[S088] Hattori, G. Hoashi, K. Matsumoto, K. and Sugaya, F. 2007. "Robust web page segmentation for mobile terminal using content-distances and page layout information". Proc. of the 16th international conference on World Wide Web (WWW '07), pp. 361–370 [WWW]

[S089] Hatzilygeroudis, I. Koutsojannis, C. and Papachristou, N. 2007. "Evaluation of Usability and Assessment Capabilities of an e-Learning System for Nursing Radiation Protection". Proc. of the Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS '07), pp. 301–306. [IEEEx]

[S090] Hijon-Neira, R. Urquiza-Fuentes, J. Dominguez-Mateos, F. and Velazquez-Iturbide, J.A. 2007. "Assessing the Usability of a Cookies-Based Access Monitoring Tool for a CMS". Proc. of the Seventh IEEE International Conference on Advanced Learning Technologies (ICALT'07), pp. 262–264. [IEEEx]

[S091] Hornbæk, K. and Frøkjær, E. 2004. "Two psychology-based usability inspection techniques studied in a diary experiment". Proc. of the third Nordic conference on Human-computer interaction (NordiCHI '04), pp. 3–12. [ACM]

[S092] Hornbæk, K. and Frøkjær, E. 2005. "Comparing usability problems and redesign proposals as input to practical systems development". Proc. of the SIGCHI conference on Human factors in computing systems (CHI '05), pp. 391–400. [ACM]

[S093] Hu, J. Zhao, J. Shima, K. Takemura, Y. and Matsumoto, K. "Comparison of Chinese and Japanese in designing B2C Web pages toward impressional usability" (2001). [IEEEx]

[S094] Hu, J. Zhao, J. Shima, K. Takemura, Y. and Matsumoto, K. 2001. "Comparison of Chinese and Japanese in designing B2C Web pages toward impressional usability". Proc. of the Second Asia-Pacific Conference on Quality Software, pp. 319–328. [IEEEx]

[S095] Hvannberg, E.T. Law, E. and Lárusdóttir, M. 2007. "Heuristic evaluation: Comparing ways of finding and reporting usability problems". Interacting with Computers, Vol. 19, Issue 2, pp. 225–240. [SD]

[S096] Iahad, N. Dafoulas, G.A. Kalaitzakis, E. and Macaulay, L.A. 2004. "Evaluation of online assessment: the role of feedback in learner-centered e-learning". Proc. of the 37th Annual Hawaii International Conference on System Sciences, pp. 10. [IEEEx]

[S097] Ivory, M.Y. and Hearst, M.A. 2001. "The state of the art in automating usability evaluation of user interfaces". Computing Surveys (CSUR) , Vol. 33, Issue 4, pp. 470–516. [ACM]

[S098] Ivory, M.Y. and Hearst, M.A. 2002. "Improving Web Site Design". IEEE Internet Computing, Vol. 6, Issue 2, pp. 56–63. [IEEE IC]

[S099] Ivory, M.Y. and Megraw, R. 2005. "Evolution of Web Site Design Patterns". Transactions on Information Systems (TOIS), Vol. 23, Issue 4, pp. 463–497. [ACM]

Appendix A (*continued*)

[S100] Jarrett, C. Quesenbery, W. Roddis, I. Allen, S. and Stirling, V. 2009. "Using Measurements from Usability Testing, Search Log Analysis and Web Traffic Analysis to Inform Development of a Complex Web Site Used for Complex Tasks". Human Centered Design, LNCS 5619, Springer, pp. 729–738. [SL]

[S101] Jati, H. and Dominic, D.D. 2009. "Quality Evaluation of E-government Website Using Web Diagnostic Tools: Asian Case". Proc. of the International Conference on Information Management and Engineering (ICIME '09), pp. 85–89. [IEEEx]

[S102] Jinling, C. and Huan, G. 2007. "Measuring Website Usability of Chinese Enterprise with a Heuristic Procedure". Proc. of the IEEE International Conference on e-Business Engineering (ICEBE'07), pp. 396–399. [IEEEx]

[S103] Johnson, T. Zhang, J. Tang, Z. Johnson, C. and Turley, J. 2004. "Assessing informatics students' satisfaction with a web-based courseware system". International Journal of Medical Informatics, Vol. 73, Issue 2, pp. 181–187. [SD]

[S104] Johnson, J. and Marshall, C. 2005. "Convergent usability evaluation: a case study from the EIRS project". Proc. of the SIGCHI Conference on Human factors in computing systems (CHI'05), pp. 1501–1504. [ACM]

[S105] Jung, H. Lee, M. Lee, S. and Sung, W. 2008. "Development of an Academic Research Information Service through Repeated Usability Evaluations". Proc. of the First International Conference on Advances in Computer–Human Interaction, pp. 195–199. [IEEEx]

[S106] Kakasevski, G. Mihajlov, M. Arsenovski, S. and Chungurski, S. 2008. "Evaluating usability in learning management system moodle". Proc of the 30th International Conference on Information Technology Interfaces (ITI'08), pp. 613–618. [IEEEx]

[S107] Kao, H.Y. 2007. "Usability Testing of AMC Hospital's Website for Home Users: Case Study for On-Site Registration Design". Proc. of the Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'07), pp. 619–622. [IEEEx]

[S108] Karahoca, D. Karahoca, A. Güngör, A. 2008. "Assessing effectiveness of the cognitive abilities and individual differences on e-learning portal usability evaluation". Proc. of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing (CompSysTech '08), article no.57. [ACM]

[S109] Katsanos, C. Tselios, N.K. and Avouris, N.M. 2006. "InfoScent evaluator: a semi-automated tool to evaluate semantic appropriateness of hyperlinks in a web site". Proc. of the 18th Australia conference on Computer–Human Interaction: Design: Activities, Artefacts and Environments (OZCHI '06), pp. 373–376 [ACM]

[S110] Kazienko, P. and Pilarczyk, M. 2006. "Hyperlink assessment based on web usage mining". Proc. of the seventeenth conference on Hypertext and hypermedia (HYPERTEXT '06), pp. 85–88. [ACM]

[S111] Keilson, S. King, E. and Sapnar, M. 1999. "Learning science by doing science on the Web". Proc. of the 29th Annual Frontiers in Education Conference (FIE '99), vol.3, pp. 13D4/7–13D412. [IEEEx]

(*continued on next page*)

Appendix A (*continued*)

[S112] Kelders, S.M. Kerkhof, S. Van Gemert-Pijnen, J.E.W. Seydel, E.R. Markus, F. and Werkman, A. 2009. "Evaluation of an Interactive Web-Based Application to Promote Healthy Behavior in Order to Maintain a Healthy Weight – Preliminary Findings". Proc. of the International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED '09), pp. 275–279. [IEEEx]

[S113] Kemp, E. and Setungamudalige, D.T. 2006. "A resource support toolkit (R-IDE): supporting the DECIDE framework". Proc. of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI (CHINZ '06), pp. 61–66. [ACM]

[S114] Kim, E. Wang, M. Lau, C. and Kim, Y. 2004. "Application and Evaluation of Personal Health Information Management System". Proc. of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS '04), vol. 2, pp. 3159–3162. [IEEEx]

[S115] Kirmani, S. 2008. "Heuristic Evaluation Quality Score (HEQS): Defining Heuristic Expertise". Journal of Usability Studies, Vol. 4, Issue 1, pp. 49–59. [JUS]

[S116] Komarkova, J. Novak, M. Bilkova, R. Visek, O. and Valenta, Z. 2007. "Usability of GeoWeb Sites: Case Study of Czech Regional Authorities Web Sites". Business Information Systems, LNCS 4439, Springer, pp. 411–423. [SL]

[S117] Koutsabasis, P. Spyrou, T. Darzentas, J. 2007. "Evaluating usability evaluation methods: criteria, method and a case study". Proc. of the 12th international conference on Human–computer interaction: interaction design and usability (HCI'07), pp. 569–578. [ACM]

[S118] Krahmer, E. and Ummelen, N. 2004. "Thinking about thinking aloud: a comparison of two verbal protocols for usability testing". IEEE Transactions on Professional Communication, Vol. 47, No. 2, pp. 105–117. [IEEEx]

[S119] Li, C. and Kit, C. 2005. "Web structure mining for usability analysis". Proc. of the IEEE/WIC/ACM International Conference on Web Intelligence, pp. 309–312. [IEEEx]

[S120] Li, Q. Sun, L. and Duan, J. 2005. "Web page viewing behavior of users: an eye-tracking study". Proc. of the International Conference on Services Systems and Services Management (ICSSSM'05), vol. 1, pp. 244–249. [IEEEx]

[S121] Lister, M. 2001. "Usability testing software for the internet". Proc of the SIGCHI conference on Human factors in computing systems (CHI '01), pp. 17–18. [ACM]

[S122] Liu, F. 2008. "Usability evaluation on websites". Proc. of the 9th International Conference on Computer-Aided Industrial Design and Conceptual Design (CAID/CD'08), pp. 141–144. [IEEEx]

[S123] López, J.M. Fajardo, I. Abascal, J. 2007. "Towards remote empirical evaluation of web pages' usability". Proc. of the 12th international conference on Human-computer interaction: interaction design and usability (HCI'07), pp. 594–603. [ACM]

[S124] Massey, A.P. Khatri, V. and Montoya-Weiss, M.M. 2008. "Online Services, Customer Characteristics and Usability Requirements". Proc. of the 41st Annual Hawaii International Conference on System Sciences (HICSS'08), pp. 33–33. [IEEEx]

[S125] Matera, M. Rizzo, F. and Carughi, G. 2006. "Web Usability: Principles and Evaluation Methods". Web Engineering, Springer, pp. 143–180. [SL]

[S126] Mayuzumi, Y. Jin, H. and Choh, I. 2003. "Study of finding solutions to problems between environmental information side and user side on Web". Proc. of the 3rd International Symposium on Environmentally Conscious Design and Inverse Manufacturing (EcoDesign'03), pp. 477–484. [IEEEx]

[S127] Milic-Frayling, N. Jones, R. Rodden, K. Smyth, G. Blackwell, A. and Sommerer, R. 2004. "Smartback: supporting users in back navigation". Proc. of the 13th international conference on World Wide Web (WWW '04), pp. 63–71. [WWW]

[S128] Millen, D.R. 1999. "Remote usability evaluation: user participation in the design of a Web-based email service". SIGGROUP Bulletin, Vol. 20, Issue 1, pp. 40–45. [ACM]

[S129] Molich, R. and Dumas, J. 2004. "Comparative usability evaluation (CUE-4)", Behaviour & Information Technology, Vol. 27, No. 3, pp. 263–281. [Other]

[S130] Molina, F. and Toval, A. 2009. "Integrating usability requirements that can be evaluated in design time into Model Driven Engineering of Web Information Systems". Advances in Engineering Software, Vol. 40, Issue 12, pp. 1306–1317. [SD]

[S131] Moraga, M.A. Calero, C. and Piattini, M. 2006. "Ontology driven definition of a usability model for second generation portals". Proc. of the 1st International Workshop on Methods, Architectures & Technologies for e-Service Engineering (MATeS 2006) in conjunction with The Sixth International Conference on Web Engineering (ICWE 2006). [ICWE]

[S132] Moreno, L. Martínez, P. and Ruiz, B. 2007. "Inclusive Usability Techniques in Requirements Analysis of Accessible Web Applications". Proc. of the International Workshop on Web Usability and Accessibility (IWWUA'07), pp. 423–428. [IWWUA]

[S133] Morgan, M. and Borns, L. 2004. "360 degrees of usability". Proc. of the SIGCHI on Human factors in computing systems (CHI '04), pp. 795–809. [ACM]

[S134] Nakamichi, N. Shima, K. Sakai, M. and Matsumoto, K. 2006. "Detecting low usability web pages using quantitative data of users' behavior". Proc. of the 28th international conference on Software engineering (ICSE '06), pp. 569–576. [ACM]

[S135] Nakamichi, N. Sakai, M. Shima, K. Hu, J. and Matsumoto, K. 2007. "WebTracer: A new web usability evaluation environment using gazing point information". Electronic Commerce Research and Applications, Vol. 6, Issue 1, pp. 63–73. [SD]

[S136] Nielsen, J. and Loranger, H. 2006. "Prioritizing Web Usability". New Riders Press, 1 edition. [Other]

[S137] Nijland, N. Seydel, E.R. van Gemert-Pijnen, J.E.W.C. Brandenburg, B. Kelders, S.M. and Will, M. 2009. "Evaluation of an Internet-Based Application for Supporting Self-Care of Patients with Diabetes Mellitus Type 2". Proc. of the International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED '09), pp. 46–51. [IEEEx]

Appendix A (*continued*)

[S138] Norman, K.L and Panizzi, E. 2006. "Levels of automation and user participation in usability testing". Interacting with Computers, Vol. 18, Issue 2, pp. 246–264. [SD]

[S139] Obendorf, H. Weinreich, H. and Hass, T. 2004. "Automatic support for web user studies with SCONE and TEA". Proc. of the SIGCHI conference on Human factors in computing systems (CHI'04), pp. 1135–1138. [ACM]

[S140] Oehler, M. and Biffignandi, S. 2008. "Web Sites Communication Performance: A Methodological Approach". Proc. of the 19th International Conference on Database and Expert Systems Application (DEXA'08), pp. 466–471. [IEEEx]

[S141] Olmsted-Hawala, E.L. Romano, J.C. and Murphy, E.D. 2009. "The use of paper-prototyping in a low-fidelity usability study". Proc. of the IEEE International Professional Communication Conference (IPCC'09), pp. 1–11. [IEEEx]

[S142] Olsina, L. Rossi, G. Garrido, A. Distante, D. and Canfora, G. 2007. "Incremental Quality Improvement in Web Applications Using Web Model Refactoring". Proc. of the International Workshop on Web Usability and Accessibility (IWWUA'07), pp. 411–422. [IWWUA]

[S143] Orii, Y. Nozawa, T. and Kondo, T. 2008. "Web-Based Intelligent Photo Browser for Flood of Personal Digital Photographs". Proc. of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT '08), Vol. 3, pp. 127–130. [IEEEx]

[S144] Oztekin, A. Nikov, A. and Zaim, S. 2009. "UWIS: An assessment methodology for usability of web-based information systems". Journal of Systems and Software, Vol. 82, Issue 12, pp. 2038–2050. [SD]

[S145] Paganelli, L. and Paterno, F. 2002. "Automatic reconstruction of the underlying interaction design of web applications". Proc. of the 14th international conference on Software engineering and knowledge engineering (SEKE' 02), pp. 439–445. [ACM]

[S146] Paganelli, L. and Paterno, F. 2002. "Intelligent analysis of user interactions with web applications". Proc. of the 7th international conference on Intelligent user interfaces (IUI '02), pp. 111–118. [ACM]

[S147] Panach, J.I. Valverde, F. and Pastor, O. 2007. "Improvement of a Web Engineering Method through Usability Patterns". Proc. of the International Workshop on Web Usability and Accessibility (IWWUA'07), pp. 441–446. [IWWUA]

[S148] Paolini, P. 1999. "Hypermedia, the Web and Usability issues". Proc. of IEEE International Conference on Multimedia Computing and Systems, Vol. 1, pp. 111–115. [IEEEx]

[S149] Pascual, V. and Dursteler, J.C. 2007. "WET: a prototype of an Exploratory Search System for Web Mining to assess Usability". Proc. of the 11th International Conference Information Visualization (IV '07), pp. 211–215. [IEEEx]

[S150] Perlman, G. 2002. "Achieving Universal Usability by Designing for Change". IEEE Internet Computing, Vol. 6, Issue 2, pp. 46–55. [IEEE IC]

[S151] Qi, Y. Reynolds, C. and Picard, R.W. 2001. "The Bayes Point Machine for computer-user frustration detection via PressureMouse". Proc. of the 2001 workshop on Perceptive user interfaces (PUI '01), pp. 1–5. [ACM]

Appendix A (*continued*)

[S152] Ramli, R. and Jaafar, A. 2008. "e-RUE: A cheap possible solution for usability evaluation". Proc. of the International Symposium on Information Technology (ITSim'08), Vol. 3, pp. 1–5. [IEEEx]

[S153] Rosenthal, A. 2007. "Redesign solution for civicinfo bc web site". Proc. of the 25th annual ACM international conference on Design of communication (SIGDOC '07), pp. 269–274. [ACM]

[S154] Roy, M.C. Dewit, O. and Aubert, B.A. 2001. "The impact of interface usability on trust in web retailers". Internet Research Journal, Vol. 11, No. 5, pp. 388–398. [IR]

[S155] Saavedra, V. Teixeira, L. Ferreira, C. and Sousa-Santos, B. 2008. "A Preliminary Usability Evaluation of Hemo@Care: A Web-Based Application for Managing Clinical Information in Hemophilia Care". Human Centered Design, LNCS 5619, Springer, pp. 785–794. [SL]

[S156] Salman, Y.B. Ince, I.F. Kim, J.Y. Cheng, H.I. and Yildirim, M.E. 2009. "Participatory design and evaluation of e-learning system for Korean language training". Proc. of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human (ICIS '09), pp. 312–319. [ACM]

[S157] Santos, B.S. Zamfir, F. Ferreira, C. Mealha, O. and Nunes, J. 2004. "Visual application for the analysis of Web-based information systems usage: a preliminary usability evaluation". Proc. of the Eighth International Conference on Information Visualisation (IV'04), pp. 812–818. [IEEEx]

[S158] Saward, G. Hall, T. and Barker, T. 2004. "Assessing usability through perceptions of information scent". Proc. of the 10th International Symposium on Software Metrics (METRIC'04), pp. 337–346. [IEEEx]

[S159] Scholtz, J. 2001. "Adaptation of traditional usability testing methods for remote testing". Proc. of the 34th Annual Hawaii International Conference on System Sciences, pp. 8. [IEEEx]

[S160] Schwerz, A.L. Morandini, M. and Da Silva, S.R. 2007. "A Task Model Proposal for Web Sites Usability Evaluation for the ErgoMonitor Environment". Human-Computer Interaction. Interaction Design and Usability, LNCS 4550, Springer, pp. 1188–1197. [SL]

[S161] Scotch, M. Parmanto, B. and Monaco, V. 2007. "Usability Evaluation of the Spatial OLAP Visualization and Analysis Tool (SOVAT)". Journal of Usability Studies, Vol. 2, Issue 2, pp. 76–95. [JUS]

[S162] Seva, R. Wu, J. and Yi, X. 2006. "Evaluation of Cinema Website". Proc. of the IEEE International Conference on Systems, Man and Cybernetics (SMC '06), Vol. 1, pp.712–717. [IEEEx]

[S163] Shanshan, Q. Buhalis, D. and Law, R. 2007. "Evaluation of the Usability of Chinese Destination Management Organisation Websites". Information and Communication Technologies in Tourism, Vol. 6, pp. 267–278. [SL]

[S164] Signore, O. 2005. "A comprehensive model for Web sites quality". Proc. of the Seventh IEEE International Symposium on Web Site Evolution (WSE'05), pp. 30–36. [IEEEx]

[S165] Skov, M.B. and Stage, J. 2005. "Supporting problem identification in usability evaluations". Proc. of the

(*continued on next page*)

Appendix A (*continued*)

17th Australia conference on Computer-Human Interaction (OZCHI '05), pp. 1–9. [ACM]
[S166] Spalteholz, L. 2008. "KeySurf: A Character Controlled Browser for People with Physical Disabilities". Proc. of the 17th international conference on World Wide Web (WWW '08), pp. 31–40. [WWW]
[S167] Sperry, R.A. and Fernandez, J.D. 2008. "Usability testing using physiological analysis". Journal of Computing Sciences in Colleges, Vol. 23, Issue 6, pp. 157–163. [ACM]
[S168] Spool, J. and Schroeder, W. 2001. "Testing web sites: five users is nowhere near enough". Proc. of the SIGCHI conference on Human factors in computing systems (CHI '01), pp. 285–286. [ACM]
[S169] Ssemugabi, S. and De Villiers, R. 2007. "A comparative study of two usability evaluation methods using a web-based e-learning application". Proc. of the annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries (SAICSIT '07), pp. 132–142. [ACM]
[S170] Stary, C. and Eberle, P. 2008. "Building up usability-engineering capability by improving access to automated usability evaluation". Interacting with Computers, Vol. 20, Issue 2, pp. 199–211. [SD]
[S171] Stefano, F. Borsci, S. and Stamerra, G. 2009. "Web usability evaluation with screen reader users: implementation of the partial concurrent thinking aloud technique". Cognitive Processing, Springer, Vol. 11, No. 3, pp. 263–272. [SL].
[S172] Stolz, C. Viermetz, M. Barth, M. and Wilde, K. 2006. "Searchstrings revealing User Intent - A better Understanding of User Perception". Proc. of the 6th International Conference on Web Engineering (ICWE'06), pp. 225–232. [ICWE]
[S173] Sulaiman, J. Zulkifli, T. Ibrahim, K.S.K. and Noor, N.K.M. 2009. "Implementing Usability Attributes In E-learning System Using Hybrid Heuristics". Proc. of the International Conference on Information and Multimedia Technology (ICIMT '09), pp. 189–193. [IEEEx]
[S174] Sullivan, T. and Matson, R. 2000. "Barriers to use: usability and content accessibility on the Web's most popular sites". Proc. of the conference on Universal Usability (CUU '00), pp. 139–144. [ACM]
[S175] Sutcliffe, A. 2002. "Assessing the reliability of heuristic evaluation for Web site attractiveness and usability". Proc. of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02), pp. 1838–1847. [IEEEx]
[S176] Tan, D. and Bishu, R. 2009. "Web evaluation: heuristic evaluation vs. user testing". International Journal of Industrial Ergonomics, Vol. 39, pp. 621–627 [Other]
[S177] Tao, Y. Guo, S. 2001. "The design of a web-based training system for simulation analysis". Proc. of the Winter Simulation Conference, pp. 645–652. [IEEEx]
[S178] Taylor, M. Wade, S. and England, D. 2003. "Informing IT system Web site design through normalization". Internet Research Journal, Vol. 13, No. 5, pp. 342–355. [IR]
[S179] Thimbleby, H. 1997. "Gentler: a tool for systematic web authoring". International Journal of Human-Computer Studies, Vol. 47, No. 1, pp. 139–168. [IJHCS]

Appendix A (*continued*)

[S180] Thompson, K.E. Rozanski, E.P. and Haake, A.R. 2004. "Here, there, anywhere: remote usability testing that works". Proc. of the 5th conference on Information technology education (CITC5 '04), pp. 132–137. [ACM]
[S181] Thompson, A.J. and Kemp, E.A. 2009. "Web 2.0: extending the framework for heuristic evaluation". Proc. of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction (CHINZ '09), pp. 29–36. [ACM]
[S182] Tirapat, T. and Achalakul, T. 2006. "Usability Assessment for Hyperlink Methods". Proc. of the International Conference on Hybrid Information Technology (ICHIT '06), Vol. 1, pp. 252–256. [IEEEx]
[S183] Toleman, M.A. and Toleman, J.M. 1998. "User experiences and a usability inspection of an electronic services environment for students". Proc. of the Australasian Computer Human Interaction Conference, pp. 87–93. [IEEEx]
[S184] Tonn-Eichstädt, H. 2006. "Measuring website usability for visually impaired people-a modified GOMS analysis". Proc. of the 8th international ACM SIGACCESS conference on Computers and accessibility (Assets '06), pp. 55–62. [ACM]
[S185] Triacca, L. Inversini, A. and Bolchini, D. 2005. "Evaluating Web usability with MiLE+". Proc. of the Seventh IEEE International Symposium on Web Site Evolution (WSE'05), pp. 22–29. [IEEEx]
[S186] Van den Haak, M.J. and De Jong, M.D.T. 2003. "Exploring two methods of usability testing: concurrent versus retrospective Think-Aloud Protocols". Proc. of the IEEE International Professional Communication Conference (IPCC'03), pp. 3. [IEEEx]
[S187] Van Velsen, L. Van der Geest, T. and Klaassen, R. 2007. "Testing the usability of a personalized system: comparing the use of interviews, questionnaires and Think-Aloud". Proc. of the IEEE International Professional Communication Conference (IPCC'07), pp. 1–8. [IEEEx]
[S188] Van Waes, L. 2000. "Thinking aloud as a method for testing the usability of Websites: the influence of task variation on the evaluation of hypertext". IEEE Transactions on Professional Communication, Vol. 43, No. 3, pp. 279–291. [IEEEx]
[S189] Vanderdonckt, J. Beirekdar, A. and Noirhomme-Fraiture, M. 2004. "Automated Evaluation of Web Usability and Accessibility by Guideline Review". Proc. of the 4th International Conference on Web Engineering (ICWE'04). pp. 28–30.[ICWE]
[S190] Vatrapu, R. and Pérez-Quiñones, M.A. 2006. "Culture and Usability Evaluation: The Effects of Culture in Structured Interviews". Journal of Usability Studies, Vol. 1, Issue 4, pp. 156–170. [JUS]
[S191] Wang, S.K. and Yang, C. 2005. "The Interface Design and the Usability Testing of a Fossilization Web-Based Learning Environment". Journal of Science Education and Technology, Vol. 14, No. 3, pp. 305–313. [SL]
[S192] Wang, L. Bretschneider, S. and Gant, J. 2005. "Evaluating Web-Based E-Government Services with a Citizen-Centric Approach". Proc. of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05), pp. 129b–129b. [IEEEx]

Appendix A (*continued*)

[S193] Wang, X. and Liu, J. 2007. "Usability Evaluation of B2C Web Site". Proc. of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCom'07), pp. 3837–3840. [IEEEx]

[S194] Weir, C. Anderson, J. and Jack, M. 2006. "On the role of metaphor and language in design of third party payments in eBanking: Usability and quality". International Journal of Human-Computer Studies, Vol. 64, Issue 8, pp. 770–784. [SD]

[S195] West, R. and Lehman, K. 2006. "Automated summative usability studies: an empirical evaluation". Proc. of the SIGCHI conference on Human Factors in computing systems (CHI '06), pp. 631–639. [ACM]

[S196] Wilkins, R. and Nyamapfene, A. 2009. "Usability driven website design – An equine sports case study". Proc. of the International Conference for Internet Technology and Secured Transactions (ICITST'09), pp. 1–6. [IEEEx]

[S197] Wilson, R. Shortreed, J. and Landoni, M. 2004. "A study into the usability of e-encyclopaedias". Proc. of the ACM symposium on Applied computing (SAC '04), pp. 1688–1692. [ACM]

[S198] Wood, F.B. Siegel, E.R. LaCroix, E.-M. Lyon, B.J. Benson, D.A. Cid, V. and Fariss, S. "A practical approach to e-government Web evaluation". IT Professional , Vol. 5, No. 3, pp. 22–28. [IEEEx]

[S199] Xu, L. Xu, B. and Jiang, J. 2005. "Testing web applications focusing on their specialties". SIGSOFT Software Engineering Notes, Vol. 30, Issue 1, pp. 10. [ACM]

[S200] Xu, L. and Xu, B. 2007. "Applying Agent into Intelligent Web Application Testing". International Conference on Cyberworlds, (CW'07), pp.61–65. [IEEEx].

Appendix A (*continued*)

[S201] Ytikseltiirk, E. 2004. "Usability evaluation of an online certificate program". Proc. of the Fifth International Conference on Information Technology Based Higher Education and Training (ITHET'04), pp. 505–509. [IEEEx]

[S202] Zaharias, P. 2006. "A usability evaluation method for e-learning: focus on motivation to learn". Proc. of the SIGCHI Conference on Human factors in computing systems (CHI '06), pp. 1571–1576. [ACM]

[S203] Zhang, P. Small, R.V. Von Dran, G.M. and Barcellos, S. 1999. "Websites that satisfy users: a theoretical framework for Web user interface design and evaluation". Proc. of the 32nd Annual Hawaii International Conference on System Sciences (HICSS'99), Vol. Track 2, pp. 8. [IEEEx]

[S204] Zhao, L. and Deek, F.P. 2006. "Exploratory inspection: a learning model for improving open source software usability". Proc. of the SIGCHI Conference on Human factors in computing systems (CHI '06), pp. 1589–1594. [ACM]

[S205] Zimmerman, D. Slater, M. and Kendall, P. 2001. "Risk communication and usability case study: implications for Web site design". Proc. of the IEEE International Professional Communication Conference (IPCC'01), pp. 445–452. [IEEEx]

[S206] Zimmerman, D. and Stapel, L. 2006. "Strategies for Integrating Research and Evaluation in Web Site Development". Proc. of the IEEE International Professional Communication Conference (IPCC'06), pp. 225–230. [IEEEx]

Appendix B. Quality assessment and data extraction form

Paper ID:	**Source:**
Evaluator:	**Data:**

Quality Assessment	**(+1)**	← → **(0)**	**(-1)**
a) The study presents a detailed description of the UEM employed.	○	○	○
b) The study provides guidelines as to how the UEM can be applied.	○	○	○
c) The study presents clear results obtained after the application of the UEM.	○	○	○
d) The study has been published in a relevant journal or conference	○	○	○
e) The study has been cited by other authors	○	○	○

Data Extraction for Sub-questions	**Answers**
Q1. Origin of the UEMs employed?	○ New ○ Existing
Q2. Underlying usability definition of UEMs employed?	○ Standard ○ Particular
Q3. Types of UEMs employed?	☐ Testing ☐ Inspection ☐ Inquiry ☐ Analytical Modeling ☐ Simulation
Q4. Type of evaluation performed by UEMs employed?	○ Manual ○ Automated
Q5. Phase(s) and Web artifacts in which the UEMs are applied?	☐ Requirements ☐ Design ☐ Implementation
Q6. Quality of the feedback provided by UEMs?	○Yes ○ No
Q7. UEMs have been empirically validated?	○ Survey ○ Case Study ○ Experiment ○ No

Notes:	

Appendix C. Mapping of the primary studies

ID	Q1		Q2		Q3					Q4		Q5			Q6		Q7				Quality score
	a	b	a	b	a	b	c	d	e	a	b	a	b	c	a	b	a	b	c	d	
S001	X			X			X			X			X	X		X	X				3.00
S002	X			X				X		X				X		X				X	2.67
S003	X			X		X				X			X	X		X	X				4.67
S004	X		X			X					X			X	X		X				2.67
S005		X	X		X					X				X		X			X		2.67
S006		X	X		X	X				X			X	X		X	X				3.00
S007		X		X	X	X	X			X			X	X		X				X	1.00
S008	X			X	X		X			X				X		X		X			1.00
S009	X			X		X					X			X		X				X	2.00
S010	X			X	X						X			X	X			X			2.33
S011	X			X		X		X			X		X	X		X				X	3.00
S012		X		X	X			X			X			X		X				X	3.00
S013		X		X	X						X			X		X		X			2.00
S014		X		X	X					X				X	X					X	0.33
S015		X		X			X			X				X		X				X	−2.00
S016		X		X		X				X			X	X	X					X	0.33
S017	X			X		X	X			X				X		X				X	1.67
S018		X		X	X	X				X			X	X		X	X				3.00
S019	X			X				X			X		X	X	X			X			4.00
S020	X			X		X				X			X	X		X				X	3.00
S021	X			X		X		X			X			X	X		X				−0.67
S022		X		X	X	X	X			X			X	X		X				X	0.00
S023	X			X		X		X	X	X				X	X				X		5.00
S024	X			X		X		X	X	X				X	X				X		5.00
S025	X			X		X		X	X		X			X		X			X		5.00
S026	X			X		X				X				X		X		X			−2.67
S027		X		X	X						X			X		X				X	1.00
S028		X		X	X	X	X			X				X	X					X	0.00
S029		X		X		X				X		X				X			X		3.33
S030	X			X		X				X			X		X				X		4.00
S031	X			X		X					X			X	X		X				2.00
S032	X		X		X		X			X				X		X				X	1.00
S033		X		X	X					X				X		X				X	−3.00
S034	X			X		X		X			X		X	X	X			X			0.00
S035		X		X	X						X			X	X					X	2.33
S036	X			X	X			X		X				X		X		X			1.33
S037		X	X				X			X				X	X		X				5.00
S038		X		X	X	X	X			X				X	X					X	1.67
S039		X		X	X		X			X				X	X					X	0.00
S040		X		X	X					X				X		X	X				−2.33
S041	X			X	X				X		X			X		X				X	0.00
S042	X			X	X				X		X			X		X		X			−0.67
S043	X			X		X				X			X	X		X			X		1.00
S044	X			X				X			X			X		X				X	0.33
S045	X			X	X						X		X	X	X					X	5.00
S046	X			X	X				X		X			X		X			X		5.00
S047		X	X		X						X			X		X				X	−0.33
S048		X	X		X			X		X				X		X				X	−2.00
S049	X		X				X	X			X			X		X			X		0.00
S050		X		X	X	X				X				X		X				X	1.33
S051		X		X		X				X				X		X				X	0.00
S052		X		X		X				X				X	X					X	2.00
S053	X		X			X				X			X			X			X		5.00
S054		X		X	X						X			X	X		X				0.00
S055		X		X	X					X				X	X			X			2.67
S056		X		X	X					X				X		X				X	2.67
S057	X			X		X				X				X	X				X		5.00
S058	X			X	X						X			X		X		X			3.00

(*continued on next page*)

Appendix c (*continued*)

ID	Q1		Q2		Q3					Q4		Q5			Q6		Q7				Quality score
	a	b	a	b	a	b	c	d	e	a	b	a	b	c	a	b	a	b	c	d	
S059	X			X	X						X			X		X				X	1.00
S060	X			X	X						X			X		X				X	1.67
S061		X		X	X	X	X	X		X		X	X	X	X					X	3.00
S062		X		X		X	X			X				X		X				X	0.00
S063		X		X	X	X				X				X		X	X				2.67
S064		X	X		X	X	X	X	X	X			X	X		X				X	2.33
S065		X		X	X					X				X		X				X	−1.00
S066		X		X	X					X				X	X					X	−0.67
S067	X		X		X						X			X		X		X			−0.33
S068		X		X	X		X			X				X		X				X	1.67
S069	X			X		X				X				X		X		X			2.67
S070		X		X			X			X				X		X				X	0.00
S071	X			X	X	X				X			X	X		X				X	2.00
S072		X	X		X		X			X				X		X				X	−3.00
S073	X			X			X			X				X		X		X			0.00
S074	X		X		X	X				X			X	X		X			X		1.00
S075		X		X	X		X			X				X	X					X	3.00
S076	X			X		X				X				X		X	X				3.00
S077	X			X	X		X			X			X		X			X			1.00
S078	X			X	X	X		X		X				X		X	X				1.67
S079		X		X				X		X			X		X					X	−0.33
S080		X	X		X	X	X			X				X	X				X		1.67
S081		X		X	X	X	X			X			X			X				X	−1.33
S082	X			X	X	X					X			X		X				X	1.33
S083		X		X	X			X		X				X	X					X	−0.67
S084		X		X			X			X				X		X	X				2.00
S085		X		X	X		X			X				X		X				X	−0.33
S086		X		X	X	X				X				X		X				X	2.00
S087		X		X	X	X	X			X				X		X				X	0.67
S088	X			X	X						X			X		X		X			−0.33
S089		X		X			X			X				X		X				X	−4.00
S090	X		X		X		X				X			X	X					X	−2.33
S091		X		X		X				X			X	X	X				X		4.00
S092		X		X	X	X				X				X	X				X		5.00
S093		X		X			X			X				X	X					X	1.67
S094		X		X	X					X				X		X				X	0.00
S095		X		X		X				X			X	X		X			X		4.33
S096		X		X			X			X				X		X	X				1.33
S097	X			X		X		X	X		X		X	X		X				X	5.00
S098		X		X	X	X	X	X	X		X		X	X		X	X				5.00
S099	X			X	X			X			X		X	X		X				X	5.00
S100		X		X	X						X			X		X				X	−2.00
S101	X		X		X						X			X		X				X	−2.00
S102		X		X		X				X				X		X				X	−1.00
S103		X		X	X	X	X			X				X		X				X	−0.33
S104		X		X		X	X			X				X		X				X	0.67
S105		X		X			X			X		X		X	X					X	1.67
S106		X	X			X	X	X		X			X	X	X					X	−1.67
S107		X	X		X		X			X				X	X					X	−0.67
S108		X	X			X	X			X				X		X				X	0.33
S109	X			X				X	X		X			X		X				X	1.33
S110		X		X	X						X			X		X			X		4.00
S111		X		X			X			X				X		X				X	1.00
S112	X			X	X		X			X				X	X					X	1.00
S113		X		X		X	X			X				X		X				X	1.67
S114		X		X	X		X				X			X	X					X	0.67
S115	X			X		X				X				X		X		X			2.00
S116	X		X			X				X			X	X	X					X	1.67
S117		X		X	X	X				X			X	X		X		X			0.00
S118		X		X	X					X				X	X				X		5.00

Appendix c (*continued*)

ID	Q1		Q2		Q3					Q4		Q5			Q6		Q7				Quality score
	a	b	a	b	a	b	c	d	e	a	b	a	b	c	a	b	a	b	c	d	
S119	X			X				X	X		X			X		X			X		2.67
S120	X			X	X						X			X		X				X	0.67
S121	X			X	X						X			X		X				X	2.00
S122		X		X	X	X	X			X			X	X		X				X	1.00
S123	X			X	X		X				X			X		X		X			1.00
S124		X		X		X	X	X		X				X		X				X	1.00
S125		X		X	X	X	X	X	X	X			X	X		X		X			5.00
S126		X		X	X	X	X			X				X		X				X	0.00
S127	X			X	X						X			X		X		X			3.00
S128		X		X	X		X			X			X	X		X				X	−1.00
S129		X		X	X		X			X				X		X		X			1.00
S130	X			X		X		X			X	X				X		X			4.00
S131	X		X			X		X		X		X	X	X	X					X	3.33
S132		X	X		X		X			X		X				X		X			1.33
S133		X		X	X					X				X	X					X	1.00
S134		X		X	X					X				X		X			X		3.00
S135	X			X	X						X			X		X		X			4.00
S136	X			X		X				X			X	X	X		X				5.00
S137		X		X	X		X			X				X		X				X	−3.00
S138		X		X	X						X			X	X					X	2.33
S139	X			X	X						X			X	X					X	3.00
S140		X		X		X				X				X		X				X	2.00
S141		X		X	X					X			X			X	X				0.00
S142	X		X					X		X			X		X			X			1.00
S143		X		X	X		X			X				X		X				X	−3.00
S144	X		X			X				X			X			X		X			4.00
S145	X			X		X		X			X			X	X					X	2.00
S146		X		X	X			X			X	X		X	X					X	4.00
S147		X	X			X		X		X			X		X			X			−1.00
S148	X			X		X				X				X		X				X	2.00
S149	X			X	X				X		X			X		X				X	0.67
S150	X			X				X		X			X		X					X	−1.00
S151	X			X	X				X		X			X		X			X		−1.67
S152		X	X		X						X			X		X				X	−0.67
S153		X		X	X	X	X			X				X	X					X	−2.33
S154		X		X			X			X				X		X				X	2.00
S155		X		X		X	X			X				X		X				X	−1.00
S156		X	X		X	X	X			X			X	X		X				X	−1.67
S157		X		X	X						X			X		X	X				2.67
S158	X		X					X		X			X	X		X			X		1.67
S159		X		X	X						X			X		X		X			2.00
S160	X		X		X			X			X			X		X				X	0.00
S161		X		X	X		X			X				X		X				X	0.00
S162		X		X		X	X			X				X	X					X	−1.00
S163		X		X			X			X				X		X				X	1.00
S164	X		X			X		X		X				X		X				X	3.00
S165		X	X		X	X	X			X				X		X			X		2.00
S166		X		X	X					X				X		X				X	0.00
S167	X			X	X		X				X			X	X		X				1.00
S168		X		X	X		X			X				X		X			X		1.67
S169		X		X		X	X			X				X		X			X		3.67
S170		X	X		X	X	X	X	X		X		X	X		X	X				−0.33
S171		X		X	X					X				X		X			X		3.00
S172	X			X	X					X				X		X		X			1.67
S173		X		X	X	X				X			X	X	X				X		2.00
S174		X		X	X						X			X		X				X	0.00
S175		X		X		X				X				X		X		X			3.67
S176		X		X	X	X				X				X		X			X		2.00
S177		X		X	X				X		X			X	X		X				−2.00

(*continued on next page*)

Appendix c (*continued*)

ID	Q1		Q2		Q3					Q4		Q5			Q6		Q7				Quality score
	a	b	a	b	a	b	c	d	e	a	b	a	b	c	a	b	a	b	c	d	
S178		X		X	X					X				X		X				X	−0.67
S179	X			X				X			X		X			X				X	1.33
S180		X		X	X						X			X		X		X			2.67
S181	X			X	X	X				X			X		X		X				2.67
S182		X	X		X		X	X		X				X	X				X		0.67
S183		X		X		X				X			X		X					X	0.00
S184		X	X					X		X			X			X				X	2.00
S185	X		X			X				X			X		X					X	4.00
S186		X		X	X					X				X		X	X				−2.00
S187		X		X	X		X			X				X		X			X		2.00
S188		X		X	X					X				X	X			X			5.00
S189	X			X		X					X			X		X				X	4.00
S190		X		X			X			X				X	X				X		1.00
S191		X		X		X				X				X		X				X	1.00
S192	X		X					X		X				X	X				X		1.00
S193		X		X	X					X				X		X				X	2.00
S194		X		X	X		X			X			X			X			X		1.33
S195	X		X		X			X			X			X		X			X		3.67
S196		X		X		X				X			X	X		X				X	1.33
S197		X		X	X		X			X				X		X				X	−1.00
S198		X		X	X	X	X			X			X	X	X					X	3.00
S199	X			X				X	X		X			X	X					X	0.33
S200		X		X	X						X			X	X					X	−1.33
S201		X		X			X			X				X		X				X	−0.33
S202		X		X			X			X				X		X	X				5.00
S203		X		X		X				X			X	X	X					X	2.67
S204	X			X		X				X				X		X				X	2.00
S205		X		X	X			X		X				X		X				X	0.33
S206		X		X	X		X			X				X		X				X	−2.00

Q1: (a) New and (b) existing.
Q2: (a) Standard and (b) ad-hoc.
Q3: (a) User testing; (b) inspection; (c) inquiry; (d) Analytical Modeling and (e) simulation.
Q4: (a) Manual and (b) automated.
Q5: (a) requirements; (b) design and (c) implementation.
Q6: (a) Yes and (b) no.
Q7: (a) Survey; (b) case study; (c) experiment and (d) no.

References

[1] S. Abrahão, E. Iborra, J. Vanderdonckt, Usability evaluation of user interfaces generated with a model-driven architecture tool, in: E. Law, E. Hvannberg, G. Cockton (Eds.), Maturing Usability: Quality in Software, Interaction and Value, Springer–Kluwer International Series in Human–Computer Interaction, Vol. 10, Springer, 2007, pp. 3–32.

[2] M. Alva, A. Martínez, J.M. Cueva, C. Sagástegui, B. López, Comparison of methods and existing tools for the measurement of usability in the Web, in: Proceedings of the 3rd International Conference on Web Engineering (ICWE'03), Spain, Springer, 2003, pp. 386–389.

[3] S. Batra, R.R. Bishu, Web usability and evaluation: issues and concerns – usability and internationalization: HCI and culture, in: Proceeding of the 2nd International Conference on Usability and Internationalization, LNCS, Vol. 4559, Springer, 2007, pp. 243–249.

[4] S.A. Becker, F.E. Mottay, A global perspective on Web site usability, IEEE Software 18 (1) (2001) 54–61.

[5] N. Bevan, International standards for usability should be more widely used, Journal of Usability Studies 4 (3) (2009) 106–113.

[6] D. Budgen, M. Turner, P. Brereton, B. Kitchenham, Using mapping studies in software engineering, in: Proceedings of PPIG 2008, Lancaster University, 2008, pp. 195–204.

[7] S. Ceri, P. Fraternali, A. Bongio, Web modeling language (WebML): a modeling language for designing Web sites, in: Proceedings of the 9th World Wide Web Conference (WWW'09), 2000, pp. 137–157.

[8] D. Cunliffe, Developing usable Web sites – a review and model, Internet Research: Electronic Networking Applications and Policy 10 (4) (2000) 295–307.

[9] N. Fenton, S.L. Pfleeger, Software Metrics: A Rigorous and Practical Approach, Second ed., International Thomson Computer Press, 1996.

[10] J.L. Fleiss, Statistical Methods for Rates and Proportions, Second ed., John Wiley & Sons, New York, 1981.

[11] A.P. Freire, R. Goularte, R.P.M. Fortes, Techniques for developing more accessible Web applications: a survey towards a process classifications, in: Proceedings of the 25th ACM International Conference on Design of communication, 2007, pp. 162–169.

[12] W. Gray, M. Salzman, Damaged merchandise? A review of experiments that compare usability evaluation methods, Human–Computer Interaction 13 (1998) 203–261.

[13] J. Gomez, C. Cachero, O. Pastor, Modeling of device-independent Web applications, IEEE MultiMedia 8 (2) (2001) 26–39.

[14] H. Hartson, T. Andre, R. Williges, Criteria for evaluating usability evaluation methods, International journal of Human–Computer Interaction 13 (4) (2001) 373–410.

[15] K. Hornbæk, Current practice in measuring usability: challenges to usability studies and research, International Journal of Human–Computer Studies 64 (2006) 79–102.

[16] E. Insfran, A. Fernández, A systematic review of usability evaluation in Web development, in: Proceedings of 2nd International Workshop on Web Usability and Accessibility (IWWUA'08), New Zealand, LNCS, Vol. 5176, Springer, 2008, pp. 81–91.

[17] International Organization for Standardization, ISO/IEC 9241-10, Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 10: Dialogue Principles, 1996.

[18] International Organization for Standardization, ISO/IEC 9241-11: Ergonomic Requirements for Office work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability, 1998.

[19] International Organization for Standardization, ISO/IEC 12207: Standard for Information Technology – Software Lifecycle Processes, 1998.

[20] International Organization for Standardization, ISO/IEC 9126-1 Standard, Software Engineering – Product Quality – Part 1: Quality Model, 2001.

[21] International Organization for Standardization, ISO/IEC 25000, Software Engineering – Software Product Quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE, 2005.

[22] M.Y. Ivory, M.A. Hearst, The state of the art in automating usability evaluation of user interfaces, ACM Computing Surveys 33 (4) (2001) 470–516.
[23] B. Kitchenham, Guidelines for Performing Systematic Literature Reviews in Software Engineering, Version 2.3, EBSE Technical Report, Keele University, UK, 2007.
[24] E. Mendes, A Systematic Review of Web Engineering Research, in: Proceedings of the International Symposium on Empirical Software Engineering (ISESE'05), 2005, pp. 498–507.
[25] C.M. Neuwirth, S.H. Regli, IEEE Internet Computing Special Issue on Usability and the Web 6 (2) (2002).
[26] J. Nielsen, in: J. Nielsen, R.L. Mack (Eds.), Heuristic Evaluation, Usability Inspection Methods, NY, 1994.
[27] J. Offutt, Quality attributes of Web software applications, in: IEEE Software: Special Issue on Software Engineering of Internet Software, 2002, pp. 25–32.
[28] K. Petersen, R. Feldt, M. Shahid, M. Mattsson, Systematic mapping studies in software engineering, in: 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), Department of Informatics, University of Bari, Italy, June 2008.
[29] J. Somervell, D. McCrickard, Comparing generic vs. specific heuristics: Illustrating a new UEM comparison technique, in: Proceedings of the Human Factors and Ergonomics Society, 2004, pp. 2480–2484.
[30] W3C: Web Content Accessibility Guidelines 2.0 (WCAG 2.0), in: Caldwell, B., Cooper, M., Guarino Reid, L., Vanderheiden, G. (Eds.), <http://www.w3.org/TR/WCAG20/>.