Presented to the Interdisciplinary Studies Program:

UNIVERSITY OF OREGON
APPLIED INFORMATION MANAGEMENT

Applied Information Management
and the Graduate School of the
University of Oregon
in partial fulfillment of the
requirement for the degree of
Master of Science

The Importance of Website Usability Testing

CAPSTONE REPORT

Julie M. Rinder
Senior Writer
Fiserv

University of Oregon
Applied Information
Management
Program

July 2012

Continuing Education
1277 University of Oregon
Eugene, OR 97403-1277
(800) 824-2714

Approved by

_____

Dr. Linda F. Ettinger
Senior Academic Director, AIM Program

The Importance of Website Usability Testing

Julie M. Rinder

Fiserv

Abstract

This scholarly annotated bibliography examines literature that investigates website usability testing. Thirty-one references consisting of peer-reviewed articles, conference proceedings, and books are reviewed. Ten usability testing categories are identified across selected testing methods including: content relevance, ease of use and decision making, graphics, interactivity between the site and the user, interface design, links, navigation/menus, page layout/text formatting, search tools, and technical performance. Most frequently listed categories address (a) navigation, (b) search features, and (c) content.

Keywords: Usability, usability testing, website usability testing, website evaluation

Table of Contents

List of Tables

Introduction

—No one ever complains that something is too easy to use—

(Tullis & Albert, 2008, Section 1.3, para. 1)

Problem

Usability is a quality that can apply to many products and services in addition to software applications and websites. For example, usability can apply to traffic signs, documents, cameras, books, tools, alarm clocks, and DVRs (Rubin & Chisnell, 2008, p. 3). What then is usability?

As defined by the International Organization for Standardization (ISO) standard 9241-11, usability is "...the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (Chiew & Salim, 2003, p. 47; Frokjaer, Hertzum, & Hornbaek, 2000, p. 345; Hornbaek & Law, 2007, p. 618; Jokela, Iivari, Matero, & Karukka, 2003, p. 53; Teo, Oh, Liu, & Wei, 2002, p. 282; Tullis & Albert, 2008, Section 1.2). Or, as Krug (2006) states, "usability really just means making sure that something works well: that a person of average (or even below average) ability and experience can use the thing—whether it's a Web site, a fighter jet, or a revolving door—for its intended purpose without getting hopelessly frustrated" (p. 5). While the term usability can apply to any product, the focus is this scholarly work is on software development and more specifically website design.

Brief history of usability. The term *usability*, when applied to software, developed out of the software industry and the work of technical communication writers. Redish (2010), explains that "...even before computer manuals had audiences other than system administrators, before online help systems, and before clear communication was built into software interfaces,

technical writers were concerned with the usability of documents" (p. 192). Human-computer interaction (HCI), the study, planning, and design of the interaction between people and computers (HCI, Wikipedia) developed in the early 1980s (Comer, et al., 1989, p. 22; Rosson, Carroll, & Rodi, 2004, p. 36). While HCI practitioners were interested in the way users interacted with the computer, user testing followed traditional laboratory based methods and "...long lists of guidelines of good practices" (Dumas, 2007, p. 55). In the late 1980s, the term "usability engineering" began to be used in the literature and practitioners conducted user testing using quantitative and practical approach (Dumas, 2007, p. 55). This work stressed the importance of the context in which the user would use the application and utilized prototypes and iterative evaluations (where the design was tested many times over the course of development) rather than a single test at the end of development (Dumas, 2007, p. 55; Lund, 2006, p. 4). In the early 1990s, Nielsen and Molich (1990) introduced a form of heuristic evaluations (i.e., the use of a list of criteria used to evaluate the software or website) that simplified heuristic evaluations and freed evaluations from the extensive guidelines used previously (Dumas, 2007, p. 55; Nielsen & Molich, 1990, pp. 249-246). Heuristic evaluation is often times called a *discount testing* method because it is less expensive than laboratory tests (Wixon, 2011, p. 201).

Over the years, additional user testing methods have been developed, including the *cognitive walkthrough* method (within which the user explains what they are thinking as they attempt to complete a task) (Dumas, 2007, p. 56). And while the methods used by usability professionals have changed, the purpose has remained the same—to create the best possible products, with the time and resources provided, that work for people (Redish, 2010, p. 196; Wixon, 2011, p. 198).

Overview of usability tests. Usability tests typically involve a user performing a task to test the ease and efficiency in which the task can be completed, as well as the user's subsequent

satisfaction with their performance or the product (Jeffries, Miller, Wharton, & Uyeda, 1991, p. 119; Usability.gov, n.d., Usability Testing). During user testing, the tester may collect both qualitative and quantitative data pertaining to the user's success, speed of performance, and satisfaction (HHS, 2006, p. 188). Usability testing helps to find problems with the user interface or design (Lynch & Horton, 2008, pp. 68–69). The following list provides brief definitions of some of the more common usability tests, as identified in the HHS *Research-Based Web Design and Usability Guidelines*, 2006, pages 196-200.

- **Automatic evaluation method**—software is used to evaluate a website and find problems, such as missing pages or links, or pages that load slowly.
- **Cognitive walkthrough**—designed to assess how well the interface supports the first-time user while they experiment to learn how to complete a task.
- **Heuristic evaluation**—usability specialist(s) examine the interface to judge how well it conforms to recognized usability principles.
- **Laboratory testing**—the user and tester are in the same location. The tester watches the user perform tasks and takes notes to report back to the development team and other interested parties.
- **Remote testing**—the user and the tester are in different locations. As an example, usability tests may be conducted via webinar with the tester watching the user use the interface.
- **Think aloud testing**—the user gives comments while performing the tasks to explain what they are thinking or why they are performing a certain activity.

Value of usability. “Usability refers to how well and how easily a user, without formal training, can interact with an information system or website [where the information system or

website is tested and not the user]" (Benbunan-Fich, 2001, p. 151) and "...the extent to which a product can be used by a specified group of users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (Whitehead, 2006, p. 788). There are many benefits to usable systems, such as increased productivity, error reduction, and reduced need for training and support (Lallemand, 2011, p. 299) and increased comprehension of the content and task necessary to complete the task (Flavian, Guinaliu, & Guerra, 2006, p. 1). For users, poor usability may result in "...an uncontrolled source of overhead caused by the need for a user to correct errors and continually relearn complex user interfaces" (NIST, 2006, The Problem). From a business perspective, operating costs are reduced by developing a product correctly the first time rather than needing to go back and refine, or totally overhaul, some aspects of the product (Usability.gov, n.d., User-centered design). As the complexity of technology grows, the need to pay close attention to usability increases (Tullis & Albert, 2008, section 1.3, para. 9). In today's business world, the uses of software are almost limitless and can be beneficial in improving productivity and streamlining processes (Al Neimat, 2005, Intro). However, if the consumer finds the product difficult to use, they will not buy the product and the business will lose money (and possibly go out of business) (Nielsen, 2000, p. 9).

Usability and websites. Usability is commonly associated with software engineering, and indeed, according to Abran, Khelifi, Suryn, and Seffah (2003), software usability may no longer be a "...luxury, but rather a basic determinant of productivity and of the acceptance of software applications" (p. 325). Usability is also associated with website engineering. As noted by Seffah and Metzker (2004), "web engineering is not a perfect clone of software engineering, but it borrows many of software engineering's fundamental concepts and principles..." (p. 74).

In the context of website design, Downing and Liu (2011), report that Nielsen claims that "...usability engineering is the key to successfully conducting commercial Web site design..."

(p. 144). Why is usability for website design so critical that Nielsen would call it a key to success? The answer involves changes to the use of the web itself. Early on, websites were what frequently amounted to electronic brochures for organizations (Nielsen, 2000, p. 15; Seffah & Metzker, 2004, p. 74). Using electronic commerce (EC), has allowed organizations to expand into previously inaccessible economic areas, reduce costs, keep inventories lean, and be more efficient and effective (Downing & Liu, 2011, p. 144; Qi, Ip, Leung, & Law, 2010, p. 78). And EC will continue to grow; electronic commerce sales in the U.S. alone are expected to increase and sales are forecasted to reach over $240 billion by 2014 (http://www.budde.com.au/Research/2010-World-Digital-Economy-E-Commerce-and-M-Commerce-Trends.html). Additionally, Internet users have increased from nearly 2 billion Internet users worldwide in June 2010 to 2.3 billion in December 2011 (www.internetworldstats.com).

Purpose

As noted by Insfran and Fernandez (2008), web applications have become the backbone of business and information interactions and the need for usability evaluation methods has become critical. "The ease or difficulty that user's experience with systems, [such as web applications], will determine their success or failure" (Insfran & Fernandez, 2008, p. 81). The purpose of this scholarly annotated bibliography is to identify literature that investigates usability testing for websites. The specific goal is to: (a) briefly describe and define usability in order to establish context and, in particular, as it applies to website development; (b) describe the rational for conducting website usability tests; and (c) list and briefly describe the key categories that are recommended in selected website usability testing methods.

Audience

The primary audience for this annotated bibliography is managers with influence over the design and content considerations of websites and website applications; the intent is to provide

information designed to help these managers to produce more usable websites (HHS, 2006, p. xv). By gaining a better understanding of the complexities involved in website usability testing, managers may be better prepared to recruit and support staff with user testing experience. A secondary audience is newcomers to the field and current website designers interested in an analysis of the key categories of one selected usability testing method in order to help them deliver more usable websites (Tullis & Albert, 2008, Preface).

Research Questions

The central research question is the broadest question the researcher can ask in the study, from which sub-questions are developed (Creswell, 2009, p. 129). The questions this annotated bibliography address are:

Main question. What is the rationale in support of website usability testing?

Sub-questions.

- What is usability testing?
- Why should websites be tested?
- What are the key categories that are recommended for inclusion in a website usability testing method?

Significance

Users use and return to websites if they can easily find useful information—in other words, the acceptability of web applications relies on their usability (Matera, Rizzo, & Carughi, 2008, p. 2). Perceived website usability is a key factor and is of critical importance in user's satisfaction with a website (Flavian, Guinaliu, & Guerrea, 2006, p. 2). One of the critical challenges facing businesses today is to develop a website that is not only compelling for the visitors, but is able to serve their business goals as well (Downing & Liu, 2011, p. 144). Website

usability testing can be used to evaluate whether websites are useful, efficient to use, and satisfying to users (Ivory & Hearst, 2001, p. 1).

Research Delimitations

Topic scope. Usability can be applied to many products and services, whether the product is a website, a treadmill, or a toaster (Tullis & Albert, 2008, Overview) or a service such as a tax preparation service. The focus of this annotated bibliography is on website usability and categories for inclusion in website usability testing.

Literature sources. Sources of literature for this annotated bibliography primarily include databases, indexes, articles, and e-collections accessible through the University of Oregon Libraries using a virtual private network (VPN) connection (use of the VPN allows access to materials offered only by subscription), such as ACM digital Library, JSTOR, EBSCO Host, and Web of Science, or through general online search engines, such as Google or Google Scholar. Additional sources of literature include the author's personal collection and academic and local public libraries. Literature from peer-reviewed academic and professional journals is preferred for inclusion in this annotated bibliography. Exceptions include works published by government or professional organizations as well as books published by authors frequently cited in peer-reviewed publications. Dissertations are generally excluded due to the variability in quality (Creswell, 2009, p. 33). Selections are evaluated based on criteria presented in the Evaluation Criteria established for this annotated bibliography.

Time frame. To allow for contributions from early literature about usability and usability testing in general, literature published between 1975 and 2012 is considered for inclusion in this annotated bibliography.

Reading and Organization Plan Preview

The following provides an overview of the plan for reading and organizing literature used in the Annotated Bibliography section of this paper. Complete details of concepts and decisions used to read and organize the literature are found in the Reading and Organization Plan description, located in the Research Parameters section of this paper.

Reading plan preview. The purpose of the reading plan is to guide the researcher in the analysis of the literature included in the Annotated Bibliography. The reading plan provides direction for the initial assessment of literature, conceptual analysis used to identify themes contained in the literature, and the final analysis of the coding results as described by Busch et al. (2005).

During the first assessment of literature, as suggested by Keshav (2007), each potential reference is initially scanned to answer the *five Cs*: category, context, correctness, contributions, and clarity (p. 1) and scanned to ensure that it addresses at least one of the research questions. During additional readings, literature is evaluated based on the literature evaluation criteria described by Bell and Smith (2009).

Using concepts of conceptual analysis, as outlined by Busch et al. (2005), to assist in analyzing the content and themes discussed and presented in the literature, each work is reviewed against a list of key terms or codes that provide context for each research question. The general guidelines to conceptual analysis, as described by Busch et al. (2005), are as follows:

1. Determine level of analysis.
2. Decide how many different concepts to code for. This step includes deciding on pre-defined words, concepts, or categories.
3. Decide whether content will be coded based on existence or frequency of a concept.
4. Decided how concepts will be distinguished from each other.
5. Develop rules for coding text.

6. Decide on a process for handling irrelevant information.
7. Actually code the texts based on the pre-defined code words and pre-determined coding rules.
8. Analyze the results.

Organization plan preview. The organization plan describes the process for sorting and presenting literature included in the Annotated Bibliography section of this paper. The references in the Annotated Bibliography section are organized by themes related to the central research questions for the study (Literature Reviews, n.d., Thematic). The intent is to categorize literature by themes that address the research questions, to aid in analysis and identification of patterns across the literature.

Definitions

The following definitions describe terminology as it is used in this annotated bibliography; definitions are excerpted from the selected literature.

Cognitive walkthrough –A usability inspection method in which the designers and developers imagine the steps performed by a user to complete a specific task and then evaluate the system responses to those tasks. After identifying tasks the user may want to perform, the participants in the evaluations typically ask four questions related to the steps they imagine the user will need to complete the task. Cognitive walkthrough can be used for a variety of products, services, software development, including website usability testing (Blackmon, Polson, Kitajima, & Lewis, 2002; Blackmon, Kitajima, & Polson, 2003; Cognitive Walkthrough, Wikipedia, n.p; Jeffries, Miller, Wharton, & Uyeda, 1991, p. 119).

Discount Usability Testing – Coined by Jakob Nielsen in 1993, refers to conducting usability tests quickly and cheaply as opposed to more expense traditional user testing performed in laboratory settings (Wixon, 2011, p. 201).

Focus Groups – "A small group of people who sit around a table and react to ideas and designs that are shown to them" (Krug, 2006, p. 133). Contrast with *Website Usability Testing*.

Heuristic – Involving or serving as an aid to learning, discovery, or problem-solving by experimental and especially trial-and-error methods <heuristic techniques> <a heuristic assumption>; also: of or relating to exploratory problem-solving techniques that utilize self-educating techniques (as the evaluation of feedback) to improve performance <a heuristic computer program> (http://www.merriam-webster.com/dictionary/heuristic).

Heuristic Evaluation – Heuristic evaluation is a discount usability engineering method for quick, cheap, and easy evaluation of a user interface design. In website usability testing,

evaluators systematically compare the website against a set list of criteria to determine how closely the design and interface follow recognized usability principles (Allen, Currie, Bakken, Patel, & Cimino, 2006, p. 413; Kjeldskov, Skov, & Stage, 2004, p. 233; Nielsen, 1992, p. 373; Nielsen, 1994, p. 152; UseIt.Com, n.d., para. 1).

Human-Computer Interaction (HCI) – Involves the study, planning, and design of the interaction between people (users) and computers. HCI is focused more on users working specifically with computers, rather than other kinds of machines (HCI, Wikipedia).

Protocol Analysis – Protocol analysis is a qualitative research method in which participants provide verbal reports about the research environment or object of research. Protocol analysis is also known as the *think aloud* method, and is useful for determining how users feel about the object of research and, in usability tests, for identifying when usability issues occur (Benbunan-Fich, 2001, p. 151; Protocol, Wikipedia).

Remote Usability Testing – A method for conducting usability test remotely, that is when the evaluators are separated from the person being tested by time or space (Castillo & Hartson, 2007, p. 2).

Think-Aloud Protocols – See *Protocol Analysis*.

Usability – The International Organization for Standardization (ISO) 9241-11 defines usability as "...the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (Jokela & Iivari, 2003, p. 53, Tullis & Albert, 2008, section 1.2; Usability.gov, n.d., Usability basics).

Krug defines usability as "...making sure that something works well [so] that a person of average (or even below average) ability and experience can use the thing...for its intended purpose without getting hopelessly frustrated" (Krug, 2006, p. 5).

Usability Evaluation – Assessing the usability of a product, in order to identify usability problems and/or to obtain measures of usability. The purpose of the evaluation can be to improve the usability or to determine whether usability objectives have been achieved (Usability BOK, n.d., Glossary).

Usability Evaluation Methods – As defined on Usability.gov (n.d), unlike usability testing, usability evaluations usually do not include a user working with the product. Instead, evaluators review the design to predict problems of successes users will have when using the website. "Evaluation techniques include: surveys/questionnaires, observational evaluations, guideline based reviews, cognitive walkthroughs, expert reviews, heuristic evaluations" (Usability.gov, n.d., Usability Evaluations).

Usability Evaluation Types – There are generally three types of usability evaluation methods: testing, inspection, and inquiry (Hom, 2001). Think-aloud protocol and remote usability testing methods are examples of the *testing* evaluation type. Focus groups and interviews are usability evaluation examples of the inquiry evaluation type. Cognitive walkthrough is an example of a evaluation method of the inspection evaluation type (Comparison of Usability Evaluation Methods, Wikipedia).

Usability Testing – "...a technique used to evaluate a product by testing it with representative users" (Usability.gov, n.d., Usability Testing). During testing, the user attempts to complete typical tasks, while observers watch, listen, take notes, and/or record the test session.

Specific to websites, Krug (2006) defines usability testing as a single person being "…shown something (whether it's a Web site, a prototype of a site, or some sketches of individual pages) and asked to either (a) figure out what it is, or (b) try to use it to do a typical task" (p. 133).

User-Centered Design (UCD) – A development methodology that involves feedback from users throughout all stages of the design and development of the product, in order to meet the user's needs. The organization's business objectives as well as the user's needs, limitations, and preferences are also considered (Usability.gov, n.d., User-Centered Design; UPA, n.d, What is User Centered Design).

Website – The following definition is from the American Heritage Dictionary (2006):

> A set of interconnected webpages, usually including a homepage, generally located on the same server, and prepared and maintained as a collection of information by a person, group, or organization.
>
> Usage Note: The transition from World Wide Web site to Web site to website as a single uncapitalized word mirrors the development of other technological expressions which have tended to take unhyphenated forms as they become more familiar. Thus email is gaining ground over the forms E-mail, especially in texts that are more technologically oriented. Similarly, there is an increasing preference for closed forms like homepage, online, and printout. (p. 1949)

The APA Style Guide (sixth edition), prefers the closed form *website* (see example section 8.03), which is the format used by this author. As discussed in sections 6.03 and 6.06 of the APA Style Guide, quotations are "reproduced word for word" (section 6.03)

directly from the material quoted. Therefore, variations of the word website (for example Web site or web site) occur in this paper.

Website Usability Testing – "The amount of success a user experiences while interacting with a website. In other words, it is the extent to which a user can achieve the desired goal during a visit to a particular website" (Atomic Design, n.d. Website Usability).

Research Parameters

This study is designed as an annotated bibliography in which information is gathered through selective reviews of pertinent literature. "An annotated bibliography serves to inform the reader of the relevance, accuracy, and quality of the sources cited" (Cornell University Library, n.d.) The following describes the search strategy used to select and review literature, including key search terms, databases searched, criteria used to evaluate literature, and a plan for reading and organizing references.

Search Strategy

The goal of the search is to identify literature that investigates usability testing of website, with an emphasis on the following: (a) selected descriptions and definitions of usability to provide context and, in particular, as it applies to website development; (b) rationale for conducting website usability testing; (c) key categories that are recommended for inclusion in a website usability testing method.

The preliminary literature search focus is on obtaining an overview of the literature available, including topics and time frames of publication, identifying key words, and identifying databases and search engines with literature pertaining to the research questions. As suggested by Creswell (2009), reference lists in scholarly articles are reviewed for citations pertinent to this study as well as additional search terms (p. 30).

Sources. Sources to search to identify literature for use in this annotated bibliography include academic journals, websites, white papers, indexes, conference papers, case studies, books, blogs, e-collections, and articles in popular magazines, with a preference for peer-reviewed literature, conference papers, and books. Literature to establish the context for usability is primarily selected from literature published between 1975 and 2012. Literature to establish a context for website usability testing and recommended categories for website testing is primarily

drawn from sources published between 1995 and 2012. Selected references published outside of those stated may be included in order to illustrate historical perspective.

Databases used. The primary search is performed using the University of Oregon Library's VPN connection to databases. Search queries are also performed through Google Scholar and Google. Searches of the ACM Digital Library, Greater Western Library Alliance University of Oregon (full text collection), return not only the search results but also a list of related journals, related magazines, related SIGs, and related conferences. Databases and search engines that may be used as part of this research, include:

- Academic Search Premier
- ACM Digital Library
- CiteSeer Scientific Literature Digital Library
- EBSCO Host
- IEEE Computer Society Digital Library (IEEE)
- Google
- Google Scholar
- JSTOR
- Safari Tech Books Online
- Web of Science
- University of Oregon library online search tools

Key terms. Preliminary key search terms are derived from the key term sections listed in peer-reviewed literature obtained from journals as well as websites of professional organizations, such as the Usability Professionals Organization (Hewitt, 1998). Additionally, a list of journals and topics covered by those journals was obtained from Emerald Publishing

(http://www.emeraldinsight.com/browse.htm?content=journal_books), to serve as a guideline for initial searches. A list of final search terms is derived from sources found through searches using the preliminary search terms. Search terms may be used in Boolean searches to specific topics. For example, website and usability or usability and web site. A list of final search terms follows.

Search terms.

- Usability
- Usability testing
- Usability testing methods
- User acceptance testing
- Website usability
- Website usability testing
- Usability evaluation
- Evaluation method

Search patterns. Searches using the preliminary search terms were conducted using various databases as described previously. Table 1 shows the results for selected search terms conducted on selected search engines. All searches were performed with University of Oregon Library's virtual private network (VPN) communication link activated, which allows access to full text sources from journals and databases in which the University subscribes.

Table 1

Preliminary Results for Selected Search Terms

Search Term	Search Engine	Results Returned
Usability	ACM (UO VPN)	21,777
	IEEE	29,838
	Google Scholar	706,000

	JSTOR	4,156
Usability testing	ACM (UO VPN)	8,453
	IEEE	230,348
	Google Scholar	178,000
	JSTOR	1,233
Usability testing methods	ACM (UO VPN)	6,949
	IEEE	230,348
	Google Scholar	139,000
	JSTOR	953
User acceptance testing	ACM (UO VPN)	3,916
	IEEE	325,227
	Google Scholar	311,000
	JSTOR	7,673
Website usability	ACM (UO VPN)	3,273
	IEEE	51,821
	Google Scholar	70,300
	JSTOR	346
Website usability testing	ACM (UO VPN)	1,592
	IEEE	243,078
	Google Scholar	32,900
	JSTOR	178

Documentation Approach

References that meet the evaluation criteria, as described in the Evaluation Criteria section following, are saved in Zotero, an online citation organization tool. Key terms, or author-defined *tag* words, are entered in Zotero to assist with later reporting and retrieval of literature. Literature that meets basic evaluation criteria, such as addressing a research question, is printed and labeled with the number assigned to the relevant research question.

Literature Evaluation Criteria

Selection of the references to be included in the Annotated Bibliography section of this paper is based on the criteria listed on the University of Oregon's library website link to *Critical Evaluation of Information Sources* located at (http://libweb.uoregon.edu/guides/findarticles/credibility.html, n.d.). The set of criteria described includes, authority, objectivity, quality, coverage, and currency.

Authority. Authority is evaluated based on considerations such as the author or publisher's credentials, affiliations, type of publisher (e.g., commercial, trade, goals of publication, etc.), reputation with peers based on citations by peers, and past publications.

Objectivity. Objectivity is evaluated based on considerations such as whether the author's goal is to "...inform, explain, educate" (Bell & Smith, 2009, Objectivity), whether the language is free of bias or emotionally charged words, whether the article cites authoritative sources, and, if applicable, whether the author's purpose seems to support the sponsoring organization's message or purpose (University of California, Berkeley Library, n.d., Objectivity).

Quality. One aspect of quality is evaluated based on the physical characteristics of the document, such as organization and format, use of correct grammar, free of spelling errors, and free of typographical errors. A second aspect of quality is based on the content of the material, such as whether facts and figures seem appropriate and agree with other experts in the field, methodologies are described, lack of questionable assumptions, and whether any facts or theories that are in disagreement with the author's thesis are explained (Bell & Smith, 2009, Quality).

Coverage. Coverage is evaluated based on whether the information presented in the literature builds on the work of other sources, includes ideas from diverse sources, or substantiates other author's work or adds new information.

Currency. Currency is evaluated based on when the literature was published or based on the timeliness of the information in respect to previous publication of similar information. As set forth in the Delimitations section of this annotated bibliography, the date for journal articles is obtained from the title page or citation information. For websites, the page creation or revision date will be used to evaluate currency. Literature published between 1990 and 2012 is given preference.

Reading and Organization Plan

Reading plan introduction. As a research method, conceptual analysis is used to study the content or themes in text (Busch et al, 2005; Insch, Moore, & Murphy, 1997) and can be used to "...identify, extract, and assess selected information from a form of media" (Bond, 2005, p. 481) in order to move from a limited set of observations to generalized inferences. Content analysis can be performed to help make sense of any recorded human communication including news media, policy documents, letters, video, novels (Baxter, 2009, p. 275; Content, n.d.; Krippendorf, 2004, p. 3), as well as "...book chapters, essays, interviews, discussions, newspaper headlines and articles, historical documents, speeches, conversations...or really any occurrence of communication language" (Busch et al., 2005, Intro to CA). The following reading plan describes the process used to read and apply content analysis techniques to code and identify themes contained in the references included in this annotated bibliography. The organization plan following describes the plan used to organize literature and present the results of the coding process.

Reading plan. As described by Holdford (2008), content analysis begins with a research question after which rules and categories by which the media is analyzed are developed (p. 175). Initially, all literature selected for use in this annotated bibliography is scanned and evaluated based on the descriptions provided in the Evaluation Criteria section of this paper (i.e., authority,

objectivity, quality, coverage and currency) as described by Bell and Smith (2009) and to determine whether the literature addresses or provides context to any of the three subtopics to the main research questions including: (a) what is usability testing, (b) why should websites be tested, and (c) what are the key categories that are recommended for inclusion in a website usability testing method. Literature that meets the basic evaluation criteria is then printed and subject to a third, more thorough deep reading as part of the content analysis research methodologies.

Content analysis. The following summarizes the eight step process suggested by Busch et al. (2005) to facilitate analysis of content and coding of texts (Steps for conducting CA), as well as the decisions used for coding the references included in the Annotated Bibliography section of this paper:

1. Determine level of analysis.

   References used in this annotated bibliography are coded based on single words, sets of words, and/or phrases, or based on themes discussed in paragraphs that relate to the words used for coding as inferred by the author (Insch, Moore, & Murphy, 1997, p. 9).

2. Decide how many different concepts to code for. This step includes deciding on pre-defined words, concepts, or categories.

   As shown in Table 2, code words and phrases are listed thematically by research question. Additional code words or phrases may be incorporated during coding to facilitate capturing important data.

3. Decide whether content will be coded based on the existence or frequency of a concept.

   References are coded based on existence of the word or concept described in Table 2, in order to derive contextual meaning.

4. Decided how concepts will be distinguished from each other.

    Concepts are coded based on whether they appear to the author to imply the same them or idea. For example, in a discussion regarding web portals, concepts related to the usability of the portal in the sense of a website are included with concepts related website usability; concepts related strictly to web portals are not coded.

5. Develop rules for coding text.

    Themes based on research questions are associated with specific codes. Concepts or themes with similar meaning are coded the same as the general term. For example, navigation system will be coded under navigation.

6. Decide on a process for handling irrelevant information.

    Irrelevant information is excluded from this study.

7. Actually code the texts based on the pre-defined code words and pre-determined coding rules.

    References are read and coded based on the identification of key words or phrases, as listed in Table 2. Searches for key words and phrases are performed manually and results are records in Zotero.

8. Analyze the results.

    The results of coding are analyzed and presented in the Conclusions section of this paper.

Table 2 lists the initial code words and phrases used to code the references included in the Annotated Bibliography, as described in step 2 above.

Table 2

Contextual Initial Code Words/Phrases by Research Question

Research Question	Code Word/Phrase
What is usability testing?	
	Usability definition
	Types of usability
Why should websites be tested?	
	Website testing
	Website evaluation methods
	Purpose of website testing
Key categories recommended for inclusion in a website usability testing method?	
	Website testing methods
	Types of website tests
	Evaluating websites

Organization plan. As discussed on the UNC Writing Center website, one way to organize a review of literature is thematically based on topics or issues found in the literature (Literature reviews, n.d, Thematic). The references included in the Annotated Bibliography section of this paper are organized into three themes, each related to a research question. Upon completion of a preliminary analysis of the title, abstract, and introduction section (if available), references are categorized by the applicable research question area, as identified based on the contextual code assigned (as shown in Table 2). The intent is to sort literature that addresses each research question area together within a section to facilitate the identification of patterns across the literature and thereby provide further understanding. References are presented in alphabetical order within each research question and literature may be listed in multiple research questions.

Usability testing. The first research question addresses usability and helps to define what usability is. References provided in this section help to provide context to the uses of usability

testing, with a special emphasis on website usability testing. These references include: (a) Bevan, Kirakowski, and Maissel, (1991), (b) Blackmon, Kitajima, and Polson (2003), and (c) Christensen and Frøkjaer (2010).

Website usability testing. The second research question explores website usability testing and why websites should be tested. References provided in this section explore website usability testing, the purposes on usability testing, and themes related to methods for conducting website usability testing. These references include: (a) Benbunan-Fich (2001), (b) Downing and Liu (2011), and (c) Fogg and Tseng (1999).

Key categories. The third research question identifies categories that are included in a selected set of website usability testing methods recommended for inclusion in a website usability testing method. References provided in this section address usability categories related to website design, function, and performance. The table format is used to summarize the categories within the selected methods. References in this section include (a) Bolchini and Garzotto (2007), (b) Chiou, Lin, and Perng (2010), and (c) Lee and Kozar (2012).

Annotated Bibliography

The annotated bibliography includes 31 references selected to address the central research question: What are the benefits of website usability testing. References are organized alphabetically by the principle author's last name within a set of categories framed by the following research sub-questions (Skidmore College, n.d.):

1. What is usability testing?
2. Why should websites be tested?
3. What are the key categories to include in a website usability testing method?

Each annotation consists of the bibliographic reference (title, author, publishers, date, etc.), the abstract, a summary of the content of each reference and how it relates to the applicable study questions, and an assessment of the credibility of the reference. Quotations provided in reference summaries are taken directly from that particular article, unless otherwise stated. Table 3 provides a list of the references included in this annotated bibliography (only the first author's name is listed for literature with more than two authors).

Table 3

Summary of References Used

Reference Author(s) (only primary author is listed for more than two authors)	**Title of Reference**
Bachiochi et al. (1997)	Usability studies and designing navigational aids for the World Wide Web
Benbunan-Fich (2001)	Using protocol analysis to evaluate the usability of a commercial web site
Bevan et al. (1991)	What is usability
Black (2002)	Usability is next to profitability
Blackmon et al. (2003)	Repairing usability problems identified by the cognitive walkthrough for the web
Bolchini & Garzotto (2007)	Quality of web usability evaluation methods: An empirical study on MiLE+

Braun et al. (2002)	Introduction: Beyond the buzz: The true meaning of usability
Chiou et al. (2010)	A strategic framework for website evaluation based on a review of the literature from 1995–2006
Christensen & Frøkjaer (2010)	Distributed usability evaluation: Enabling large-scale usability evaluation with user-controlled instrumentation
Davis & Shipman (2011)	Learning usability assessment models for web sites
Downing & Liu (2011)	Assessing web site usability in retail electronic commerce
Fogg & Tseng (1999)	The elements of computer credibility
Hallahan (2001)	Improving public relations web sites through usability research
Hartson et al. (1996)	Remote evaluation: The network as an extension of the usability laboratory
HHS—U.S. Dept. of Health and Human Services (2006)	Research-based web design & usability guidelines
Hollingsed & Novick (2007)	Usability inspection methods after 15 years of research
Krug (2006)	Don't make me think: Common sense approach to web usability
Lee & Kozar (2012)	Understanding of website usability: Specifying and measuring constructs and their relationships
Levi & Conrad (2008)	Usability testing of world wide web sites
Lynch & Horton (2008)	Web style guide: Basic principles for creating web sites
Manzari & Trinidad-Christensen (2006)	User-centered design of a web site for library and information science students: Heuristic evaluation and usability testing
Nielsen (1994)	Enhancing the explanatory power of usability heuristics
Nielsen (2000)	Designing web usability: The practice of simplicity
Nielsen & Molich (1990)	Heuristic evaluation of user interfaces
Olmsted-Hawala et al. (2010)	Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination websites for usability
Palmquist (2001)	An overview of usability
Qi et al. (2010)	A new framework on website evaluation
Redish et al. (2002)	Usability in practice: formative usability evaluations - evolution and revolution
Spool (2009)	The $300 million button
Tedeschi (1999)	E-commerce report: Online merchants find that a well-designed web site can have a big impact on bottom line
Whitehead (2006)	Evaluating web page and web site usability

What is Usability Testing?

To address this research question, references are included that provide definitions or describe usability, usability testing in general, as well as references that provide definitions and descriptions of specific website usability testing methods.

Bevan, N., Kirakowski, J., & Maissel, J. (1991). What is usability. *Proceedings of the 4th International Conference on HCI*. Retrieved from http://www.usabilitynet.org/papers/whatis92.pdf

Abstract. The paper relates different approaches to usability based on the product, the user, ease-of-use, actual usage and the context of use; and proposes that usability should be defined as the ease of use and acceptability of a product for a particular class of users carrying out specific tasks in a specific environment. Criterion levels for measurements of attitude and user performance determine whether the design of the product is successful in achieving usability. Diagnostic evaluation of usability problems may be based on analysis of user interaction or comparison of product attributes with guidelines.

Credibility. Nigel Bevan received a PhD in Man-Machine Interaction (MMI) as well as degrees in physics and psychology. Dr. Bevans has authored 80 publications, has edited or contributed to several international standards, and was the Director of Professional Development for the Usability Professionals Association. Dr. Jurek Kirakowski is a senior lecturer at University College Cork and Director of the Human Factors Research group. Jonathan Maissel is vice president of product at Tok Media. This article was published in the Proceedings of the fourth International Conference on Human-Computer Interaction.

Summary. This article provides a definition of usability and describes some of the methods of evaluating usability and techniques for specifying usability. Early definitions

of usability were derived from the views of what usability was and how it should be measured. From a product oriented viewpoint, ergonomic attributes about the product were measured. From a user-oriented viewpoint, the user's mental effort and attitude were measured. From a user performance viewpoint, how easy the product is to use or whether the product would be used in the real world were measured (p. 1). Later definitions of usability encompass all three views and are defined as "...the ease of use and acceptability of a system or product for a particular class of users carrying out specific tasks in a specific environment" (p. 2). The product attributes determine the ease of use and are measured by user performance and satisfaction and determines whether a product will be used and how it will be used (p. 2). Writing about measuring usability, the authors note the users testing to determine ease of use can focus on performance in laboratory settings or testing in the real world. Some of the measures of performance include goal achievement, work rate, knowledge acquisition, and operability (p. 3). The internal state of the user can also be measured to assess usability, for example, by measuring muscular or ocular fatigue, affective state, and mental state. User testing can be used to compare the product or system with other products or systems to establish goals for effectiveness, efficiency, and satisfaction.

Blackmon, M., Kitajima, M., & Polson, P. G. (2003). Repairing usability problems identified by the cognitive walkthrough for the web. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 497–504). doi:10.1145/642611.642698 http://doi.acm.org/10.1145/642611.642698

Abstract. Methods for identifying usability problems in web page designs should ideally also provide practical methods for repairing the problems found. Blackmon et al. [2] proved the usefulness of the Cognitive Walkthrough for the Web (CWW) for identifying

three types of problems that interfere with users' navigation and information search tasks. Extending that work, this paper reports a series of two experiments that develop and prove the effectiveness of both full-scale and quick-fix CWW repair methods. CWW repairs, like CWW problem identification, use Latent Semantic Analysis (LSA) to objectively estimate the degree of semantic similarity (information scent) between representative user goal statements (100-200 words) and heading/link texts on each web page. In addition to proving the effectiveness of CWW repairs, the experiments reported here replicate CWW predictions that users will face serious difficulties if web developers fail to repair the usability problems that CWW identifies in web page designs.

Credibility. Marilyn Blackmon has a PhD in Cognitive Psychology from the University of Colorado, Boulder, where she is also a research associate and lecturer. Dr. Blackmon has published 17 articles in peer-reviewed journals and has presented at various conferences (http://autocww.colorado.edu/~blackmon/Papers/BlackmonMHCurriculumVitae111017.pdf). Muneo Kitajima, PhD is a professor in the Management and Information Systems Engineering department at Nagaoka University of Technology in Japan, has published 10 journal articles and 73 conferences papers, co-written four book chapters (http://kjs.nagaokaut.ac.jp/mkitajima/index.html). Peter Polson received a PhD from the University of Indiana and is Professor Emeritus of Psychology and a Faculty Fellow of the Institute of Cognitive Science at the University of Colorado, Boulder (http://psych.colorado.edu/~ppolson/).

Summary. This article provides a definition of a specific type of website usability test, *cognitive walkthrough for the web*. The authors define cognitive walkthrough for the web (CWW) as "...a theoretically-based usability inspection method for detecting and

correcting design errors that interfere with finding information on a website" (p. 497). While CWW is based on cognitive-walkthrough methods used in general interface design testing, "CWW is specially tailored to simulate users navigating a website and better fits a realistic website design process" (p. 497). In this method, the design team imagines the user performing a task and asks the following questions about the user's ability to complete the identified task: (a) will the user try to achieve the right effect, (b) will the correct action be evident to the user, (c1) will the user connect the correct sub-region of the page based on heading information, (c2) will the user understand which widget (i.e., button, menu, etc.) to use based on labels and descriptions, and (d) will the user interpret the system's response correctly (p. 498). The authors note that questions a, b, and d are common to the original cognitive walkthrough method, but questions c1 and c2 are "...specifically adapted to the user's web navigation process..." (p. 498). The authors conclude that through the process of applying CWW methods, "the design team can successfully repair usability problems for users *highly similar* to themselves in background knowledge" (p. 503). Using CWW methods in conjunction with other techniques, such as latent semantic analysis (a method of identifying semantic relatedness of terms (p. 498)), designers can more readily identify and apply terminology that is comprehensible to a greater number of users (p. 504).

Braun, K., Gadney, M., Haughey, M., Roselli, A., Synstelien, D., Walter, T., & Wertheimer, D. (2002). Introduction: Beyond the buzz: The true meaning of usability. In M. Holzschlad & B. Lawson (Eds.), *Usability: The site speaks for itself* (pp. 2–13). Acocks Green, Birmingham, UK: Glasshaus.

Abstract. This book is about web usability of the sites from the designer of those sites. The authors discuss their initial designs, their audiences, how they got feedback on the

sites, how they made design tweaks to meet the unique needs of that group of users. This book is for designers, web professionals, and site owners who want to consider a variety of options when planning the future sites, and for everyone who is interested in the future, and the pioneers of this revolutionary new medium.

Credibility. The introduction is written by Molly Holzschag, a senior web accessibility strategist at Knowability, Inc. Her clients include AOL, Adobe Systems, eBay, Microsoft, and Yahoo. Ms. Holzschag received a Master of Arts in Media Studies from New School University and is a steering committee member of the Web Standards Project (WaSP). Ms. Hozschag has published 35 books related to web design.

Summary. In the introduction to the book *Usability: The site speaks for itself*, editor Molly Hozschag introduces and discusses usability in general and then provides an over view of usability testing. The author explains that just as many other products go through website usability testing, so do websites. The author notes that website are tested to ensure that they are in fact useful and safe. After noting that "entire processes for testing the usability of websites exist, with little standardization" (p. 5), the author then describes the following general steps for usability testing (see page 6):

- User needs are researched.
- User needs are identified and documented.
- Basic prototypes, such as wireframes, are developed and tested. The basic prototypes include text and structural components.
- The site is built.
- Focus groups of potential users are gathered.

- Usability engineers watch the focus group participants as they use the site, by using video cameras, one way screens, and other observational tools.
- Focus group members are encouraged to speak aloud as they move through the site.
- Results are compiled and changes are made to the site as necessary.

The author then describes some of the views and approaches of pundits of website usability testing, including a summary of Jakob Nielsen's ideas. The author recognizes the usefulness of website usability testing approaches and practices recommended by Nielsen, but believes that a one-size-fits-all approach is short-sighted and does not take into account "...something as amorphous as website design..." (p. 6). Because this book is written for not only web designers, but lay-people interested in learning more about web design and usability testing, the audience of this annotated bibliography may find this introduction, indeed the entire book, entertaining and informational.

Christensen, L., & Frøkjaer, E. (2010). Distributed usability evaluation: Enabling large-scale usability evaluation with user-controlled instrumentation. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries,* 118–127. doi: 10.1145/1868914.1868932 http://doi.acm.org/10.1145/1868914.1868932

Abstract. We present DUE (Distributed Usability Evaluation), a technique for collecting and evaluating usability data. The DUE infrastructure involves a client-server network. A client-based tool resides on the workstation of each user, providing a screen video recording, microphone input of voice commentary, and a window for a severity rating. The idea is for the user to work naturalistically, clicking a button when a usability problem or point of uncertainty is encountered, to describe it verbally along with illustrating it on screen, and to rate its severity. These incidents are accumulated on a server, providing access to an evaluator (usability expert) and to product developers or

managers who want to review the incidents and analyze them. DUE supports evaluation in the development stages from running prototypes and onwards. A case study of the use of DUE in a corporate environment is presented. The study indicates that the DUE technique is effective in terms of low bias, high efficiency, and clear communication of usability issues among users, evaluators and developers. Further, DUE is supporting long-term evaluations making possible empirical studies of learnability.

Credibility. Erik Frøkjær is a computer scientist and associate professor in the Computer Department at the University of Copenhagen, who joined the university following 12 years of employment in the private sector. Mr. Frøkjær has published 19 articles. This article is published in the proceedings of the 6th Nordic Conference on Human-Computer Interaction. This article is well organized and presents the background, purpose, and research methods in an easy to understand manner.

Summary. This paper provides a definition of a specific type of website usability test, the Distributed Usability Evaluation (DUE). As a website evaluation method, DUE uses an instrumentation framework, consisting of an instrumentation client and an instrumentation server (p. 120). In software, instrumentation refers to code that is inserted in an application to record the values or function parameters, timing statistics, or other information (Cole, 2009, p. 3). As a website evaluation tool, DUE software is loaded on the client (user's) machine. When the user wants to record a usability issue, they activate the DUE instrumentation software which then makes a record of the usability issue, including a "...recorded video of the user's screen, the recording of the user's explanation, sound from surroundings, plus the collection of timestamps, and severity ratings the user reported [as] issue(s)" (p. 120). The recordings are periodically transmitted over the internet to a server location. Evaluators, developers, and development managers can later access and analyze

the data. The authors' list of advantages include that the developers do not build anything special in their applications, DUE is easy to learn and any user can quickly record usability issues while doing their actual work rather than in artificial laboratory settings, developers can see exactly what the user saw and experienced, and that many users can report findings, as opposed to some usability evaluation techniques, which use a limited number of evaluators. At the conclusion of testing, the authors note that the usability issues found related primarily to efficiency and satisfaction aspects of usability issues (p. 126). Therefore, "this indicates that DUE evaluations should be combined with evaluation techniques directed at identifying effectiveness issues" (p. 120).

Hallahan, K. (2001). Improving public relations web sites through usability research. *Public Relations Review, 27*(2), 223–239. doi:10.1016/S0363-8111(01)00082-0

Abstract. Public relations communicators can use a variety of techniques to pretest the effectiveness of messages, including readability testing, focus groups, peer and jury reviews, experiments and field testing/test marketing. Another yet-untapped technique is usability research, a set of procedures used for more than 50 years in product and software development. Although usability is a technique that can be applied to any type of public relations message, usability research has particular relevance for enhancing the effectiveness of websites and other new interactive techniques. This article examines the nature and value of usability research, and the elements of an effective website based on usability principles. Applications to other types of public relations communications are also discussed.

Credibility. At the time of publication, Kirk Hallahan, now a full professor (see biography at http://lamar.colostate.edu/~hallahan/), was an associate professor in the Department of Journalism and Technical Communication at Colorado State University.

Improving Public Relations Web Sites Through Usability Research is published in the peer-reviewed Public Relations Review and has been cited at least 55 times by other authors. The references cited in this article are from journals, including the *Journal of Public Relations* and the *Public Relations Review*, reputable authors, such a Jakob Nielsen and Jared Spool, or books.

Summary. Although the phrase *usability research* is used in the title and abstract for this article, the author equates usability research to testing by using phrases such as "test[ing] military hardware" (p. 224) or testing applied to human-computer interfaces (HCIs), such as computer hardware and software (p. 224). The author defines websites as a specific type of HCI. Hallahan cites definitions of usability, such as "... how well the intended users can interact with technology to carry out an assigned activity" (p. 224), or "an interface that is workable and intuitive from the user's point of view" (p. 224), and goes on to say that "usability research strives to improve both the efficiency and effectiveness of systems" (p. 225). Hallahan categorizes usability testing into two broad categories of (1) laboratory testing and (2) observation and expert assessments conducted outside the laboratory (p. 225), and describes techniques used in both of these categories. In the category of *laboratory testing*, Hallahan explains the process of using a laboratory setting to test, equipment used for testing, such as video cameras and two-way mirrors, session structure and length, the value of laboratory testing, and differences between focus groups and user testing. A table summarizing the procedures used for conducting usability tests in a laboratory setting is included (p. 227). In the category for *other observation and assessment techniques*, Hallahan identifies website inspection and user inquiry testing methods. Website inspection involves expert assessments of websites, where the testers typically use lists or guidelines (also known as heuristic evaluation) to

test the website (p. 228). Inquiry-based research involves directly observing the user and asking questions (p. 228). Alternatively, surveys and polls can be used the solicit feedback from the user.

Hartson, H., Castillo, J., Kelso, J., Kamler, J., & Neale, W. (1996). Remote evaluation: The network as an extension of the usability laboratory. *Proceedings of the SIGCHI conference on Human factors in computing systems: Common ground*. (pp. 228–235). doi:10.1145/238386.238511 http://doi.acm.org/10.1145/238386.238511

Abstract. Traditional user interface evaluation usually is conducted in a laboratory where users are observed directly by evaluators. However, the remote and distributed location of users on the network precludes the opportunity for direct observation in usability testing. Further, the network itself and the remote work setting have become intrinsic parts of usage patterns, difficult to reproduce in a laboratory setting, and developers often have limited access to representative users for usability testing in the laboratory. In all of these cases, the cost of transporting users or developers to remote locations can be prohibitive. These barriers have led us to consider methods for remote usability evaluation wherein the evaluator, performing observation and analysis, is separated in space and/or time from the user. The network itself serves as a bridge to take interface evaluation to a broad range of networked users, in their natural setting.

Credibility. H. Rex Hartson is currently Emeritus Professor of Computer Science at Virginia Tech. He received Bachelor of Science and Master of Science degrees in electrical engineering (EE) from the University of Michigan in 1965 and 1967 and Master of Science and PhD from Ohio State University in 1972 and 1975. Dr. Hartson has published over 66 articles since 1969 and is a founding faculty member in Human-Computer Interaction in the department of Science at Virginia Tech

(http://www.cs.vt.edu/user/127). At the time of publication, José Castillo was a Master's degree candidate at Virginia Polytechnic Institute and is attributed with five peer-reviewed articles since 1996. John Kelso has a Master's degree in Computer Science from George Washington University has published it least five peer-reviewed articles since 1995. Wayne C. Neale has a PhD in Industrial and Systems Engineering and is currently Vice President Product Management at MyWorld, Inc. (http://www.linkedin.com/in/wayneneale). This article is published in the peer-reviewed proceedings of the ACM Special Interest Group (SIG) for Computer-Human Interaction (CHI) Conference on Human Factors in Computing Systems.

Summary. This article provides a definition of a specific type of usability testing technique—remote evaluation—and summarizes various methods used to conduct remote evaluation. The authors define "... remote evaluation to be usability evaluation wherein the evaluator, performing observation and analysis, is separated in space and/or time" (p. 228). While traditional usability evaluation is conducted in laboratories, this method can be limited by the expense of equipping and maintaining the lab and having access to evaluators. Transporting evaluators to the lab or usability experts and developers to the evaluators can be expensive. Additionally, in some cases, testing networks and remote work settings may be part of the usage pattern being tested. Remote evaluation techniques can help to fill this gap. The following list briefly describes some of the possible approaches to remote evaluation (the authors note that this is not necessarily a complete list).

- Portable evaluation—the laboratory is taken to the users in their natural work setting.
- Local evaluation at remote site—an evaluation lab located near the target evaluators is contracted to conduct usability testing. Design documents, software, samples, and

prototypes are delivered to the local lab. The quality of this service can vary, based on the evaluation methods used. This evaluation method may not be suitable for all developments groups (p. 229).

- Remote inspection—developers send their designs to remote contractors who then evaluate the design based on user profiles, intuitive inspection (where the evaluator makes judgments based on their experience or prior knowledge), design guidelines, and software standards. Results may vary based on the knowledge and skills of providers (p. 229).
- Remote questionnaire/survey—questionnaires or surveys display during use of the application to ask users questions about their usage. While data is captured while reactions are fresh with the users, data in limited to the questions presented, which may not address topics of interest to the evaluators (p. 229).
- Remote-controlled evaluation—the remote evaluator stays at their location and connects to the computer in the usability lab and performs tasks. Audio to hear what the evaluator is saying can be recorded and video camera or other screeb capture software and track and record the user's actions. Equipment must be configured to allow this kind of interaction, but testing can be arranged at times convenient for both parties, and frequently at less expense than transporting evaluators and/or testers (p. 229).
- Instrumented remove evaluation—an application is embedded with code to capture user actions, including keystrokes or mouse clicks, executed during task performance. Journal and logs are later analyzed for usability issues. It may be difficult to indentify true usability problems based on information in the logs (p. 230).

- Semi-instrumented remove evaluation—users are trained to identify usage events that negatively or positively impact their task performance or satisfaction and transmit information about these events to developers. Context information about the system, tasks, and interface history is included in the feedback transmission (p. 230).

HHS—U.S. Dept. of Health and Human Services. (2006). *Research-based web design & usability guidelines*. US Government Printing Office. Retrieved from http://www.usability.gov/guidelines/guidelines_book.pdf

Abstract. The Research-Based Web Design and Usability Guidelines (*Guidelines*) were developed by the U.S. Department of Health and Human Services (HHS), in partnership with the U.S. General Services Administration. The Guidelines were developed to assist those involved in the creation of Web sites to base their decisions on the most current and best available evidence. The Guidelines are particularly relevant to the design of information-oriented sites, but can be applied across a wide spectrum of Web sites.

Credibility. The Guidelines are developed and published by U.S. government departments. Contributors consist of various experts, including some listed in this annotated bibliography, including Robert Bailey, PhD, Joseph Dumas, PhD, Melody Ivory, PhD, and Janice Redish, PhD. Each Guideline is accompanied by a list of sources (references), which are listed in the Sources section at the end of the book. The sources primarily consist of literature published in peer-reviewed journals, books, or conference proceedings.

Summary. The authors describe and define website usability testing. Usability testing is performed "...to identify issues that keep users from meeting the usability goals of a Web site" (p. 203). Usability testing methods include automated evaluations, inspection evaluations, operational evaluations, and human performance testing (p. 203).

Hollingsed, T., & Novick, D. G. (2007). Usability Inspection methods after 15 years of research. *SIGDOC ’07 Proceedings of the 25th annual ACM international conference on design of communication* (pp. 668–668). doi:10.1145/1297144.1297200

Abstract. Usability inspection methods, such as heuristic evaluation, the cognitive walkthrough, formal usability inspections, and the pluralistic usability walkthrough, were introduced fifteen years ago. Since then, these methods, analyses of their comparative effectiveness, and their use have evolved in different ways. In this paper, we track the fortunes of the methods and analyses, looking at which led to use and to further research, and which led to relative methodological dead ends. Heuristic evaluation and the cognitive walkthrough appear to be the most actively used and researched techniques. The pluralistic walkthrough remains a recognized technique, although not the subject of significant further study. Formal usability inspections appear to have been incorporated into other techniques or largely abandoned in practice. We conclude with lessons for practitioners and suggestions for future research.

Credibility. Tasha Hollingsed, now Associate Chief Engineer at Lockheed Martin, holds a Master's of science degree in computer science from the University of Texas at El Paso and was a senior software engineer at the time this article was published. As listed by ACM, Hollingsed has published two articles and has been cited eight times. "David G. Novick, AT&T Distinguished Professor in Engineering earned his J.D.at Harvard University in 1977 and his PhD in Computer and Information Science at the University of Oregon in 1988. Before coming to UTEP he was on the faculty of the Department of Computer Science and Engineering at the Oregon Graduate Institute and then Director of Research at the European Institute of Cognitive Sciences and Engineering"

(http://www.cs.utep.edu/DeptCS/people/facultyList.html#pteller). Dr. Novick has published over 56 articles and has been cited 150 times.

Summary. This article describes five types of usability testing and the conclusions regarding the use of the methods "...and which [evaluation methods] led to relative methodological dead ends" (p. 249). For each usability method, the authors provide a description of the method and how the testing is conducted, a brief history of the methodology, advantages and disadvantages, and a summary of the usefulness and how the testing is used today. The following provides a brief description of each testing method:

- Heuristic evaluation—compared to empirical testing (which finds more severe issues, but at greater cost), heuristic evaluation can find many problems, but is cheaper to conduct. In heuristic evaluation, professional usability experts or amateurs evaluate the interface for problems using a defined set of guidelines. This method is still widely used and research continues to explore ways to apply heuristic evaluation techniques to a variety of interfaces and projects (p. 250).
- Cognitive walkthrough—this method is characterized by evaluating the interface for ease of use and learning, based on cognitive models of learning and use (p. 250). The following four steps are used during the analysis phase: 1. A goal is set for the user to complete; 2. The available actions are determined; 3. The user selects the action that they think will take them closer to their goal; 4. The user performs the action and evaluates the feedback given by the system (p. 250). Cognitive walkthrough is used with semantic analysis to evaluate websites and is used to identify three types of problem areas: domain vocabulary, knowledge, meaningful sub-regions and link

labels, and conventions used to represent page elements (p. 251). This method is still widely in use today.

- Pluralistic usability walkthrough—this evaluation methodology uses many of the same techniques as cognitive walkthrough with the exception that users are made up of groups, usually product team members or usability experts, who act together as a single user. To take full advantage of this methodology, evaluators should approach the system with different intents of focus. The authors note that "...it seems likely that use of the pluralistic usability walkthrough is widespread but teams do not refer to it as such in published reports. Rather the pluralistic feature of the walkthrough may have become such a standard practice that it need not be mentioned" (p. 251).
- Formal usability inspection methods—in this evaluation method, human factors experts review the interface to look for potential user problems. The evaluators use a variety of techniques, including task performance models, heuristics, and other human factors expertise (p. 251). This method gains speed in testing at the expense of the perspective of multiple stakeholders. Due to lack of recent literature , the authors were unable to conclude whether this technique is commonly in use today.

Krug, S. (2006*). Don't make me think: Common sense approach to web usability* (2nd ed.). Berkeley, Ca: New Riders Press.

Abstract. Five years and more than 100,000 copies after it was first published, it's hard to imagine anyone working in Web design who hasn't read Steve Krug's "instant classic" on Web usability, but people are still discovering it every day. In this second edition, Steve adds three new chapters in the same style as the original: wry and entertaining, yet loaded with insights and practical advice for novice and veteran alike. Don't be surprised

if it completely changes the way you think about Web design (http://www.peachpit.com/store/product.aspx?isbn=0132809893).

Credibility. As posted on www.howto.gov (2011), a website maintained and promoted by the U.S. General Services Administration, Steve Krug has worked for 15 years as a usability expert for companies such as Apple, Bloomberg.com, Lexus.com, and NPR, to help make products and websites that people could use and enjoy. His book Don't Make Me Think has sold over 300,000 copies. Krug is the CEO of his consulting firm, Advanced Common Sense, and spends most of his time teaching usability workshops and consulting (http://www.howto.gov/training/classes/essentials-of-usability-testing). Krug makes references to authors such as Jared Spool, Janice Redish, and Jakob Nielsen. The audience for this annotated bibliography may find this book informative and easy to read and a good background to the components of websites.

Summary. Written for the lay audience, *Don't Make Me Think*, describes usability testing, how users really use the website, categories to test, provides suggestions for working with managers not familiar with usability testing and websites, and ends with a list of recommended reading. Although information about usability testing is interspersed throughout the book, chapter nine specifically defines usability testing and explains why usability testing is important. Krug clarifies the differences between focus groups and usability tests, and defines usability tests as "...one user at a time is shown something (whether it's a Web site, a prototype of a site, or some sketches of individual pages) and asked to either (a) figure out what it is, or (b) try to use it to do a typical task" (p. 133). Krug provides sections on various aspects of usability testing, including definitions and examples of how to complete the tests. Examples include how many users to use, how to work with the test participants, who should do the testing, who should observe, and what

to test. Examples of the pros and cons of traditional usability tests and discount usability testing methods, what Krug calls *Lost-our-Lease Testing*, are provided. The following are some of Krug's key statements about website usability testing: (a) testing one user is 100% better than testing none; (b) testing one user early in the project is better than testing 50 near the end; (c) the importance of recruiting representative users is overrated; (d) the point of testing is not to prove or disprove something. It's to inform your judgment; (e) testing is an iterative process; and (f) nothing beats a live audience reaction.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *CHI '90 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People*, 249–256. doi:10.1145/97243.9728

Abstract. Heuristic evaluation is an informal method of usability analysis where a number of evaluators are presented with an interface design and asked to comment on it. Four experiments showed that individual evaluators were mostly quite bad at doing such heuristic evaluations and that they only found between 20 and 51% of the usability problems in the interfaces they evaluated. On the other hand, we could aggregate the evaluations from several evaluators to a single evaluation and such aggregates do rather well, even when they consist of only three to five people.

Credibility. Nielsen holds a PhD in Human-Computer Interaction from the Technical University of Denmark in Copenhagen. Nielsen has written 12 books and authored many articles. Rolf Molich holds a Master's in software engineering from the Technical University of Denmark, has worked as a principle investigator at the Nielsen Norman Group, and currently owns and manages DialogDesign. Nielsen and Rolf are co-inventors of the heuristic evaluation method (http://www.upassoc.org/upa_

publications/jus/2007august/useful-usable.pdf). This article has been cited by others over 1354 times.

Summary. This article provides a definition and description for conducting one type of usability test—*heuristic evaluation.* The authors explain that heuristic evaluation is testing that is performed by looking at an interface and forming an opinion of what is good and bad about the interface (p. 249). In previous publications, the authors had narrowed existing heuristic guidelines down to the following nine principles: (a) simple and natural dialogue; (b) speak the user's language; (c) minimize memory load; (d) be consistent; (e) provide feedback; (f) provide clearly marked exits; (g) provide shortcuts; (h) good error messages; (i) prevent errors. After conducting a series of experiments, the authors conclude that heuristic evaluation is a valid testing technique, but because it can be difficult for one tester to find all the faults, the heuristic evaluation should be performed by approximately five individuals. The authors conclude that more than five reviewers does not provide a significantly greater benefit. The major benefits of heuristic evaluation are that:

- It is low cost to conduct
- Is intuitive to learn
- Does not require advanced planning
- Can be done early in the developmental process

A disadvantage is that it may identify usability problems without providing suggestions for fixing the problem, and the results of testing may be biased by the evaluators and, as a result, may not lead to breakthroughs in the design process (p. 255).

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010). Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-

dissemination websites for usability. *Proceedings of the 28th International Conference on Human Factors in Computing Systems*. 2381–2390. Retrieved from http://dmtlarchive.skku.edu/share/CHI2010%20Proccedings%20%20/p2381.pdf

Abstract. We describe an empirical, between-subjects study on the use of think-aloud protocols in usability testing of a federal data-dissemination Web site. This double-blind study used three different types of think-aloud protocols: a traditional protocol, a speech-communication protocol, and a coaching protocol. A silent condition served as the control. Eighty participants were recruited and randomly pre-assigned to one of four conditions. Accuracy and efficiency measures were collected, and participants rated their subjective satisfaction with the site. Results show that accuracy is significantly higher in the coaching condition than in the other conditions. The traditional protocol and the speech communication protocol are not statistically different from each other with regard to accuracy. Participants in the coaching condition are more satisfied with the Web site than participants in the traditional or speech-communication condition. In addition, there are no significant differences with respect to efficiency (time-on-task). This paper concludes with recommendations for usability practitioners.

Credibility. All authors were current or retired researchers for the US Census Bureau at the time of publication. Erica Olmsted-Hawala holds an MA in technical communication and is a user experience researcher at the U.S. Census Bureau. Olmsted-Hawala publishes in government research publications and international conference proceedings. Elizabeth Murphy holds a PhD in cognitive psychology. Murphy has 30 years of experience in user-interface design and evaluation (http://www.upassoc.org/upa_publications/jus/2011november/romano-bergstrom-authors.html). Sam Hawala is a statistician for the U.S. Census Bureau and has published

over 10 articles. Kathleen Ashenfelter holds a PhD in Quantitative Psychology from the University of Notre Dame and is a research psychologist with the Human Factors and Usability Group at the U.S. Census Bureau. This article was published in the Proceedings of the 28th International Conference on Human Factors in Computing Systems.

Summary. In this article, the authors discuss one type of usability testing method—the *think-aloud protocol*—and define three techniques of using the think-aloud protocol to conduct website usability tests. Think-aloud protocol, also known as protocol analysis (Benbunan-Fich, 2011), refers to the user verbalizing their thoughts as they use, typically be completing pre-assigned tasks, on a website (p. 2381). The three think-aloud protocols techniques reviewed are the (a) traditional, (b) speech-communication, and (c) coaching techniques. The following summarizes and briefly defines these techniques during website usability tests:

- Traditional—the test administrator attempts to keep the participant talking by using non-intrusive words or phrases, including "Keep talking" or "Um-humm," (p. 2382), and avoids asking the user why they are doing something.
- Speech-communication—in this technique, because "speech-communication theory holds that the ways human beings naturally communicate within a speaker/listener relationship include a certain amount of acknowledgement and feedback..." (p. 2383), the test administrator communicates with the participant more than that allowed in the traditional technique and uses words and phrases, including um-hum, oh, and okay, to acknowledge the user.
- Coaching—in this technique, the test administrator encourages and may help or assist the participant. The test administrator may also ask direct questions about the website,

such as how the participant likes a color or what the participant thinks of a certain button.

The authors conclude by noting that the results of usability tests conducted using think-aloud protocols may be skewed based on the testing technique. While participants tested using a coaching method may experience a higher success rate with completing tasks, "this is detrimental in a typical usability study because the coaching injects bias into the results: the results are skewed toward better performance than the participants would have achieved without help" (p. 2388).

Palmquist, R.A. (2001). An overview of usability. *Journal of Education for Library and Information Scienc*e, *42*(2), pp. 123-136. Retrieved from http://www.jstor.org/stable/40324025

Abstract. This article provides an overview of usability literature and testing methods. Usability is a term shared by a wide variety of interests in the commercial software development arena, and it is a notion that argues for the centrality of the user—a focus long held by the library and information science (LIS) professions. As more information centers and libraries turn to a Web-based portal to introduce their users to the resources and services they provide, the need to create usable Web-based information displays emerges as a critical task for LIS professionals. Usability testing can be helpful in capturing the user's experience with electronically provided information and can be helpful in determining the success of Web-based efforts. This study was initially supported by a research grant proposal award from the Association for Library and Information Science (ALISE) 1996-97.

Credibility. Ruth Palmquist received a PhD in Information Transfer from Syracuse University and has taught at the university level since 1988. Dr. Palmquist is an assistant

professor at the University of Texas at Austin and has authored 17 articles published in peer-reviewed journals. This article is published in a peer-reviewed journal.

Summary. The author defines usability and website usability testing, reasons for performing website usability testing, and methods used for website usability testing. Three approaches to usability are described: (a) usability engineering, sometimes used to refer to the user involvement during the entire development process; (b) usability analysis or inspection, which refers to user testing during the developmental process to fine-tune designs; and (c) usability testing, which refers to testing completed, or nearly complete products (p. 126). However, the author notes that these distinctions tend to become blurred and in actual practice usability testing tends to depend on the company culture (p. 126). When conducting website usability testing, user tasks should be defined. "The tasks the user is asked to perform determine, to a large degree, whether you study what the user does using an existing interface or whether you observe testing of particular features of the interface itself" (p. 131). The author then describes some of the techniques and methods used for testing the website, including: (a) usability testing on the completed (or nearly completed) site to evaluate the user's experience and functionality of the website; (b) heuristic evaluations using a list of guidelines; (c) think-aloud protocols, which help to identify user's misconceptions about the website design; and (d) the browse and bookmark test, where the user bookmarks different link that think may be useful to address a particular task, which is used to assess what the user thinks is useful (p. 131). The author concludes by noting that "...the effort to provide some degree of user testing can produce better Web-design efforts....which are a high priority for many in the e-commerce and for-profit business community....[and] nonprofit sector as well" (p. 132).

Why Should Websites be Tested?

This section includes references that describe the importance, rationale, or value of performing usability testing in general and/or website usability testing in particular.

Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management*, *39*(2), 151–163. doi:10.1016/S0378-7206(01)00085-4

Abstract. Despite the increasing popularity of electronic commerce, there appears to be little evidence of the methodical evaluation of the usability of commercial web sites. The usability of a web site defines how well and how easily a visitor, without formal training, can interact with the site. This paper reports the results of a research project, which applies a systematic qualitative technique known as protocol analysis or think aloud method, to examine the usability of a commercial web site. About 15 usability principles and 3 evaluation parameters (content, navigation and interactivity) were used as a framework to analyze the verbal protocols of a sample of users interacting with a greeting card web site. The protocols provided evidence of usability problems caused by crowded content, poor navigation and cumbersome interactivity. These results underscore the importance of two crucial usability goals for commercial web sites: clear path to products and transparency of the ordering process.

Credibility. Raquel Benbunan-Fich received a PhD in Management Information Systems from Rutgers University – Graduate School of Management and is currently Associate Professor of Information Systems at Baruch College. Dr. Benbunan-Fich has published over 43 journal articles (refereed), authored numerous book chapters, and presented over 32 times at academic conferences. At this time of publication, Benbunan-Fich was

Assistant Professor of Information Systems at the Stillman School of Business at Seton Hall University. This article is published in a refereed journal.

Summary. This article first provides a limited review of literature related to commercial websites and summarizes the importance of usability testing for commercial websites. The author notes that "…better usability will result in more efficient interaction between user and the site and will increase the probability that the user will return and/or make a purchase" (p. 151) and "the easier the navigation through the product choices, the better the chances that users [will] buy something" (p. 160). Because the use of commercial websites is discretionary and training users prior to their use of the website in not an option, the author states that "…usability testing is of the upmost importance. Better usability will result in more successful interactions between the user and the site and may help the user to make a purchase" (p. 161).

Black, J. (2002). Usability is next to profitability. December 4, 2002, special report. *Bloomberg Business Week Online*. New York, NY. Retrieved online on May 27, 2012, from http://www.businessweek.com/technology/content/dec2002/tc2002124_2181.htm

Abstract. Software companies such as Oracle, Apple, and Microsoft have benefited from usability practices and testing. As software from progressed from being offered via expensive mainframe computer, accessible by few, to being accessed by many via personal computers, the need to ensure that even grandma can operate the software hidden behind the front end of a website is even more critical. Companies lacking easy to learn software lose sales.

Credibility. At the time of publication, Jane Black was a reporter covering new technology. She has written for publications including Boston Magazine and the Washington Post, and covered technology and international news for the BBC and Business Week Online.

According to Bloomberg's website (http://www.bloomberg.com/company/#history), over 440 publications use Bloomberg's content and over 26 million visit Bloomberg news sites monthly.

Summary. This article explains that by employing usability practices, such as usability engineering and usability testing, businesses can improve profits and market share. As businesses demand products that are easy to learn and use, more and more businesses, including Oracle, Apple, and Microsoft, employ usability experts and usability testing. In the 1970s, when computer software ran over expensive main frame computers, the idea of allocating time and money to "...the task of designing screens that simplify and speed up the use of a particular program or even an entire computer network was laughable." Today, however, as software is accessed via personal computer and more and more users, "...it became economically sensible to spend time developing resource-intensive user interfaces...." The use of websites to access complicated backend software has made usability more critical. Black reports that Jakob Nielsen has stated that e-commerce sites lose nearly half of their potential sales because visitors can't figure out how to use them. The author provides examples of businesses who have improved their profitability by employing usability practices, such as Dell Computer which saw sales improve by $33 million per day following the implementation of usability practices to their e-commerce website.

Downing, **C. E., & Liu, C.** (2011). Assessing web site usability in retail electronic commerce. *2011 IEEE 35th annual computer software and applications conference (COMPSAC)* (pp. 144–151). doi:10.1109/COMPSAC.2011.26 http://dx.doi.org/10.1109/COMPSAC.2011.26

Abstract. In the increasingly competitive environment of electronic commerce, companies are paying careful attention to Web site design and function to attract and

retain both traffic and customers. One key factor that has been shown to increase both is Web site usability. This paper presents and tests a Web site usability research framework derived from prior literature. Fourteen Fortune 500 retail Web sites are examined by 261 potential customers and rated on aspects of usability. Results show that content and ease of use are givens with regard to usability, but identity, download delay, trust assurance, made-for-the-medium, responsiveness and emotion can all be differentiators.

Credibility. Charles E. Downing received a PhD from Northwestern University and is the assistance professor of Management Information in the Wallace E. Carroll School of management at Boston College. Chang Liu received a Doctor of Business Administration (DBA) from Mississippi State University where he majored in Management Information Systems. Dr. Liu also has Master's degrees in business economics and management information systems. Dr. Liu is a professor and operations management and information systems (OM&IS) chair at Northern Illinois University College of Business. This article is published in the Proceedings of the 2011 35th IEEE Annual Computer Software and Applications conference.

Summary. The authors describe the economic environment related to websites, define website usability, and describe the use of usability testing to evaluate the websites of 14 large business organization. The authors describe the growth of e-commerce and an increase in the use of high-speed internet by the world population and U.S. population, and notes, "companies need to ensure that their Web presence is accessible and acceptable to this huge market to maintain future profitability" (p. 144). Part of being accessible and acceptable to consumers is to not only have a web presence, but to have a usable website. The authors believe that developing a website that is not only compelling for the visitors but also meets business goals is one of the critical challenges businesses

face today (p. 144). The authors summarize previous research and research and the factors identified by which websites may be assessed. These factors include: (a) identity or purpose of the website, (b) downloading delay, (c) content, (d) ease of use, (e) trust assurance, (f) made-for-the-medium (design that differentiates product and service offerings), (g) responsiveness, (h) promotion of the business, and (i) emotion (the users emotional response to the website) (pp. 145–146). User testing was used to evaluate the several factors related to website usability, including (a) ease of use, (b) content, (c) download delay, (d) emotion, and (e) over all sense of website usability. Following the testing, the authors conclude that of the factors listed, promotion of the business is not important, assuming that the website is generally usable. Users expect websites to be easy to use and have quality content. Additionally, special interest should be given to factors such as download speed, trust, responsiveness, and emotion, as "...such attention could make the difference for a web site in the increasingly competitive world of electronic commerce" (p. 149).

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. *Proceedings of the SIGCHI conference on human factors in computing systems: The CHI is the limit* (pp. 80–87). doi:10.1145/302979.303001 http://doi.acm.org/10.1145/302979.303001

Abstract. Given the importance of credibility in computing products, the research on computer credibility is relatively small. To enhance knowledge about computers and credibility, we define key terms relating to computer credibility, synthesize the literature in this domain, and propose three new conceptual frameworks for better understanding the elements of computer credibility. To promote further research, we then offer two perspectives on what computer users evaluate when assessing credibility. We conclude

by presenting a set of credibility-related terms that can serve in future research and evaluation endeavors.

Credibility. BJ Fogg, PhD, is the director of the Persuasive Tech Lab at Stanford University. Dr. Fogg received his PhD in psychology with a focus on user experience and has published 25 articles in peer-reviewed journals and one book. This article was published in the 1999 proceedings of the SIGCHI conference on Human Factor in computing systems conference and has been cited 51 times.

Summary. In this article the authors discuss credibility and how it relates to computer products and note that "as a community of HCI [human-computer interaction] professionals, we should be concerned about the products we create, research, and evaluate" (p. 80). The authors define credibility as *believability* and explain that based on research, most researchers identify trustworthiness and expertise as the important key components of credibility (p. 80). Trustworthiness relates to the user's perception of the goodness or morality of the source. Expertise relates to the user's perception of the knowledge and skill of the source (p. 80). The authors continue by listing eight instances when credibility seems to matter (they acknowledge the list is not comprehensive) when applied to computer products. They are, when computer products: (a) act as knowledge sources, (b) instruct or tutor users, (c) act as decision aids, (d) report measurements, (e) run simulations, (f) render virtual environments, (g) report on work performed, (h) report about their own stats. In discussing research regarding credibility applied to computer products, the authors cite research showing that "computers gain credibility when they provide information that users find accurate or correct" (p. 81), that "small computer errors have a disproportionally large effects on perceptions of credibility" (p. 82), and that "once users perceive that a computer product lacks credibility, they are likely to stop

using it, which provides no opportunity for the product to regain credibility" (p. 82). Some researchers have focused on variables that affect credibility for computer products including content and interface design [a component of website design] (p. 82). The authors discuss "For example, a web page may appear credible just because of its visual design" (p. 83). A list of factors users evaluate when assessing credibility is provided, including (a) interface, (b) functional, and (c) information credibility, and the authors mention that web site evaluators should look at all three of these aspects. Finally, the authors conclude that providing credibility in computer products is important because "...people may not naturally separate the credibility of one aspect of a computer product to another" (p. 85). Likewise, perceptions about one aspect of the computer product affect their perception of the entire product (p. 85).

Hallahan, K. (2001). Improving public relations web sites through usability research. *Public Relations Review*, *27*(2), 223–239. doi:10.1016/S0363-8111(01)00082-0

Abstract. Public relations communicators can use a variety of techniques to pretest the effectiveness of messages, including readability testing, focus groups, peer and jury reviews, experiments and field testing/test marketing. Another yet-untapped technique is usability research, a set of procedures used for more than 50 years in product and software development. Although usability is a technique that can be applied to any type of public relations message, usability research has particular relevance for enhancing the effectiveness of websites and other new interactive techniques. This article examines the nature and value of usability research, and the elements of an effective website based on usability principles. Applications to other types of public relations communications are also discussed.

Credibility. At the time of publication, Kirk Hallahan, now a full professor (see biography at http://lamar.colostate.edu/~hallahan/), was an associate professor in the Department of Journalism and Technical Communication at Colorado State University. *Improving Public Relations Web Sites Through Usability Research* is published in the peer-reviewed Public Relations Review and has been cited at least 55 times by other authors. The references cited in this article are from journals, including the Journal of Public Relations and the Public Relations Review, reputable authors, such a Jakob Nielsen and Jared Spool, or books.

Summary. Although the phrase *usability research* is used in the title and abstract for this article, the author equates usability research to testing by using phrases such as "test[ing] military hardware" (p. 224) or testing applied to human-computer interfaces (HCIs), such as computer hardware and software (p. 224). The author defines websites as a specific type of HCI. Hallahan cites definitions of usability, such as "... how well the intended users can interact with technology to carry out an assigned activity" (p. 224), or "an interface that is workable and intuitive from the user's point of view" (p. 224), and goes on to say that "usability research strives to improve both the efficiency and effectiveness of systems" (p. 225). Hallahan provides the following reasons for conducting website testing: (a) to ensure ease of use and satisfaction by users; (b) establishment of a benchmark for future versions; (c) to minimize service, training, and support costs; (d) increased use and support of the product (or site) by users; (e) the imperative to be competitive (p. 230). Additionally, statistics and monetary amounts are applied to the cost of poor website usability versus the cost of usability testing (p. 231).

HHS—U.S. Dept. of Health and Human Services. (2006). *Research-based web design & usability guidelines*. US Government Printing Office. Retrieved from http://www.usability.gov/guidelines/guidelines_book.pdf

Abstract. The Research-Based Web Design and Usability Guidelines (*Guidelines*) were developed by the U.S. Department of Health and Human Services (HHS), in partnership with the U.S. General Services Administration. The Guidelines were developed to assist those involved in the creation of Web sites to base their decisions on the most current and best available evidence. The Guidelines are particularly relevant to the design of information-oriented sites, but can be applied across a wide spectrum of Web sites.

Credibility. The Guidelines are developed and published by U.S. government departments. Contributors consist of various experts, including some listed in this annotated bibliography, including Robert Bailey, PhD, Joseph Dumas, PhD, Melody Ivory, PhD, and Janice Redish, PhD. Each Guideline is accompanied by a list of sources (references), which are listed in the Sources section at the end of the book. The sources primarily consist of literature published in peer-reviewed journals, books, or conference proceedings.

Summary. To answer the question of why websites should be tested, the authors note that "....users define usability as their perception of how consistent, efficient, productive, organized, easy to use, intuitive, and straightforward it is to accomplish tasks within a system" (p. 3). In order to ensure that websites meet user expectations and provide the best possible outcomes, the authors recommend that the appropriate usability tests are conducted frequently during the design process, and the findings are used to make changes to the website (p. 1). "Generally, the more iterations [of testing and updating the design], the better the Web site" (p. 188).

Krug, S. (2006*). Don't make me think: Common sense approach to web usability* (2nd ed.). Berkeley, Ca. New Riders Press.

Abstract. Five years and more than 100,000 copies after it was first published, it's hard to imagine anyone working in Web design who hasn't read Steve Krug's "instant classic" on Web usability, but people are still discovering it every day. In this second edition, Steve adds three new chapters in the same style as the original: wry and entertaining, yet loaded with insights and practical advice for novice and veteran alike. Don't be surprised if it completely changes the way you think about Web design (http://www.peachpit.com/store/product.aspx?isbn=0132809893).

Credibility. As posted on www.howto.gov (2011), a website maintained and promoted by the U.S. General Services Administration, Steve Krug has worked for 15 years as a usability expert for companies such as Apple, Bloomberg.com, Lexus.com, and NPR, to help make products and websites that people could use and enjoy. His book *Don't Make Me Think* has sold over 300,000 copies. Krug is the CEO of his consulting firm, Advanced Common Sense, and spends most of his time teaching usability workshops and consulting (http://www.howto.gov/training/classes/essentials-of-usability-testing). Krug makes references to authors such as Jared Spool, Janice Redish, and Jakob Nielsen. The audience for this annotated bibliography may find this book informative and easy to read and a good background to the components of websites.

Summary. Written for the lay audience, *Don't Make Me Think*, describes usability testing, how users really use the website, categories to test, provides suggestions for working with managers not familiar with usability testing and websites, and ends with a list of recommended reading. Just as information about usability testing is interspersed throughout the book, so are reasons for conducting website usability testing. The following describe

some of the highlights to why websites should be tested. In chapter two, Krug describes that users, happy to finally found some information about the desired topic may stay and muddle through a poorly designed website rather than spending time and energy to find a different website (p. 28). Developers and designers may ask why the design matters if users are going to muddle through it anyway (p. 28). Krug states, "the answer is that it matters a great deal because while muddling through may work sometimes, it tends to be inefficient and error-prone" (p. 28). Krug continues that if users feel like they "get the site," they are more likely to find what they are looking for, look at additional content on the site, and be repeat visitors to the site (p. 29). In chapter 10, Krug explains that the theory of goodwill towards a website and how if that goodwill diminishes due to poor design [and by extension usability], the user's good will to the website and even the company may diminish. By depleting the user's goodwill toward the business and the website, the company's "...brand—which they [in this case an airline] spend hundreds of dollars a year polishing—had definitely lost some of its luster" (p. 162), to the user. Krug then lists things that diminish the user's goodwill toward the company, including hiding sought for information, requiring particular formatting, for example dashes in Social Security numbers, and asking for unnecessary information (pp. 164–165). Things that increase goodwill include (a) making the main information users want readily available; (b) identifying extra expenses, such as shipping costs; (c) streamlining steps; and (d) answer the questions the user is likely to have (pp. 166–167).

Levi, M.D., & Conrad, F.G. (2008). Usability testing of world wide web sites. *U.S. Bureau of Labor Statistics Office of Survey Methods Research*. Retrieved from http://www.bls.gov/ore/htm_papers/st960150.htm

Abstract. Building a medium or large World Wide Web site, whether for distribution over the Internet or over an intranet, can and should be viewed as a major software development effort. Once Web site creation is seen as software development, it becomes natural to apply the tools and methods we have learned in past projects. The life cycle of Web creation is identical to that of traditional software: requirements gathering, analysis, design, implementation, testing, and deployment. And, just as traditional software development should have a functionality and a usability component, so should Web development efforts. But, just as comprehensive functional requirements and a detailed design document do not by themselves guarantee that a programmer's final code will be correct, so up-front usability guidelines do not by themselves guarantee a usable end product. In both cases a distinct validation process is required. This article describes Web site usability testing methods and their application in the development of Web sites.

Credibility. Michael Levi received a Master's degree in Public Administration from the Kennedy School of Government at Harvard University and has served as Associate Commissioner for Publications and Special Studies at the Bureau of Labor and Statistics (BLS) since 2007. Mr. Levi initiated the use of the Internet at BLS, for communicating with customers (http://www.bls.gov/bls/senior_staff/levi.htm). Frederick G. Conrad received a PhD in Cognitive Psychology from the University of Chicago and is an associate of research scientist in the Institute of Social Research at the University of Michigan and a research associate professor in the Joint Program of Survey Methodology at the University of Maryland. At the time of publication, Dr. Conrad authored two books and published 30 articles in peer-reviewed journals. This article is published by the U.S. Bureau of Labor Statistics Office of Survey Methods.

Summary. The authors describe website usability testing, why website usability should conducted, and then explain criteria that should be tested during usability testing. Usability is defined as "…the degree to which a given piece of software assists the person sitting at the keyboard to accomplish a task, as opposed to becoming an additional impediment to such accomplishment" (Background). The criteria that are typically accessed during usability testing include: (a) ease of learning, (b) retention of learning time, (c) speed of task completion, (d) error rate, and (e) subjective user satisfaction. Additionally, systems are measured and weaknesses are identified through usability testing (Background). Usability is necessary, because, "…just as comprehensive functional requirements and a detailed design document do not by themselves guarantee that a programmer's final code will be correct, so up-front usability guidelines do not by themselves guarantee a usable end product. In both cases, a distinct validation process is required" (Background). The three main styles of usability testing include (a) *exploratory testing* (areas that confuse users, cause slow down, or mistakes), (b) *threshold testing* (system performance is measured against predetermined goals), and (c) *comparison testing* (different approaches are compared to determine which best suits the user's needs (Background). Methods used for testing include card sorting, heuristic evaluation, scenario-based testing, and questionnaires. As concluding comments, the authors state, "there is no question in our minds that our systems are better because of the usability testing we have performed, and that the end users have benefited in direct, measurable ways" (Conclusion).

Nielsen, J. (2000). *Designing* w*eb usability: The practice of simplicity*. Indianapolis, IN: New Riders Publishing.

Abstract. *Designing Web Usability* is the definitive guide to usability from Jakob Nielsen, the world's leading authority. Over 250,000 Internet professionals around the world have turned to this landmark book, in which Nielsen shares the full weight of his wisdom and experience. From content and page design to designing for ease of navigation and users with disabilities, he delivers complete direction on how to connect with any web user, in any situation. Nielsen has arrived at a series of principles that work in support of his findings. This book is a must-have for anyone who thinks seriously about the web (http://www.peachpit.com/store/product.aspx?isbn=156205810X).

Credibility. Nielsen holds a PhD in Human-Computer Interaction from the Technical University of Denmark in Copenhagen. Nielsen has published 12 books and authored numerous articles.

Summary. In this book, Dr. Nielsen provides an explanation of what usability testing is, why websites should be tested and categories to test, by explaining good web design practices and providing examples on nearly every page. The author states that the information contained in "this book is based on observations of usability tests with about 400 users from a wide variety of backgrounds using a large number of different websites over the last six years" (p.14). In describing the importance of website usability tests, the author states the following:

> Usability has grown dramatically in importance for web-based companies because of an inversion in the relationship between user experience and the ability to separate customers from their money. In the old world that is populated by most computer companies, customer would pay for a product first and only later take it home and discover you need a two-inch-thick book to figure out how to format chapter headings. (p. 388)

Today, "usability rules the web....if the customer can't find a product, then he or she will not buy it" (p. 9). By extension, website usability testing helps to ensure that websites are usable and that the user can readily find what they are looking for. On page 15, the author provides a high-level list of fundamental errors made when designing websites. Four major categories listed include: (a) page design, (b) content design, (c) site design, and finally (d) intranet design. Within each chapter, the author provides additional details regarding each of the four major design categories, instructions based on previous user testing, examples of evidence from user testing, as well as screen shots showing iterative examples of design changes. The book concludes with chapters on accessibility, international use, future predictions, and a conclusion.

Redish, J., Bias, R.G., Bailey, R., Molich, R., Dumas, J., & Spool, J. (2002). Usability in practice: Formative usability evaluations - evolution and revolution. In *CHI '02 extended abstracts on Human factors in computing systems* (*CHI EA '02)*. ACM, New York, NY, USA, doi:10.1145/506443.506647 http://doi.acm.org/10.1145/506443.506647

Abstract. Formative evaluation is a collection of "find-and-fix" usability engineering methods, focused on identifying usability problems before a product is completed. In this forum, four experienced usability professionals will address different aspects of formative evaluations.

Credibility. Ginny Redish has a PhD in Linguistics from Harvard University, has published over 30 articles in peer-reviewed journals, and has authored or co-authored three books. Randolph Bias received a PhD from the University of Texas at Austin, where he is currently an Associate Professor. Dr. Bias "…has worked in industry for over 20 years as a usability engineer…" including work at Bell Labs, IBM, and BMC software (http://www.ischool.utexas.edu/people/person_details.php?PersonID=52). Rolf Molich

holds a Master's in software engineering from the Technical University of Denmark, has worked as a principle investigator at the Nielsen Norman Group, and currently owns and manages DialogDesign. Joe Dumas received a PhD in Cognitive Psychology from SUNY Buffalo, and has published over 11 articles in peer-reviewed journals or books. Jared Spool is the founder of User Interface Engineering, the largest usability research organization of its kind in the world (as described on the UIE website), has published over 26 articles or books, and has been cited 155 times. Robert Bailey received a PhD in Human Performance form Rice University is currently president of Computer Psychology, Inc., and has worked in the Human Performance Technology Center at Bell Laboratories and the Human Engineering Laboratory at Lockheed Aircraft Corporation in Burbank, California. Dr. Bailey has published over 200 articles and technical reports and has published three books (http://www.linkedin.com/pub/bob-bailey/1a/118/612).

Summary. This article describes how usability testing relates to usability engineering and the reason for using usability tests. The authors introduce the concept of formative evaluation, "...user testing with the goal of learning about the design to improve its next iteration" (p. 885), and then discuss the use of formative evaluation techniques. The authors describe the website design process and four categories of website usability methods: (a) automated evaluation, (b) inspection evaluations, (c) operational evaluations, and (d) performance testing. Bailey describes how each method is used at different times during the design process, but cautions that evaluators must perform analysis to determine whether the identified usability problem is a true usability problem. The authors note that usability tests and evaluations are used to help design usable websites (p. 885), and that usability tests, and subsequent reports of usability tests can be used to identify the most important usability issues, identify positive findings, classify the

seriousness of identified issues, and report how many participants encountered each problem. As the authors note, "the primary purpose of the usability test is to cause beneficial improvements to the user interface..." (p. 887). The authors provide techniques for using usability testing in the work environment to help build collaboration among team members, and the importance of using and documenting usability techniques, methods, and results to avoid similar problems in the future (p. 889).

Spool, J., (2009). The $300 million button. *User Interface Engineering*. Retrieved from http://www.uie.com/articles/three_hund_million_button/

Abstract. It is hard to imagine that a form could be simpler: two buttons, and one link. Yet it turns out that this form was costing a business $3000,000,000 a year. The form appeared as part of the online purchase process after shoppers had already selected items to purchase. The intent of the form was to register users, so that the shoppers could check out more quickly on subsequent visits. Shoppers resented having to register in order to purchase items. The form intended to make shopping easier, turned out to only help a small percentage of the customers. Usability testing would have identified this problem.

Credibility. Jared Spool is the founder of User Interface Engineering, the largest usability research organization of its kind in the world (as described on the UIE website), has published over 26 articles or books, and has been cited 155 times.

Summary. Spool explains how performing usability tests revealed an issue that cost the company $300,000,000 in sales a year (para. 1). Once the issue was fixed, online sales increased an extra $15 million the first month (para. 14). During usability testing, researchers observed that users were reluctant to complete a registration form prior to completing their purchases. Rather than register, many users left the site and did not complete their purchase. Later analysis revealed that for those users who did register,

45% had multiple registrations on the system (para. 11). Spool notes that prior to the usability test, "…the designers of the site had no clue there was even a problem" (para. 1). When the designing the registration process, the design team did not think that anyone would mind registering, because it would save them [the users] time in the long run. They were wrong about the first time shoppers (para. 6). Once usability testing was performed, the issue was fixed (by including a simple message that registration was not required to complete the purchase) and sales increased dramatically.

Tedeschi, B. (1999, August 30). E-commerce report: Online merchants find that a well-designed web site can have a big impact on bottom line. *New York Times*. Retrieved May 30, 2012, from http://www.nytimes.com/1999/08/30/business/e-commerce-report-line-merchants-find-that-well-designed-web-site-can-have-big.html?pagewanted=print&src=pm

Abstract. This articles describes IBM's efforts to redesign their website. The newly designed website saw a reduction in the use of the search tool (a positive indicator that users were able to find what they are looking for) and sales increased 400%. The project included several components, including information architecture, navigation, graphic design, and the selection of the words and photographs for each page.

Credibility. Bob Tedeschi is a graduate of the Columbia Journalism School, teaches writing at his local community colleges, and has been published in a variety of publications, including Wired magazine, Yankee magazine, and the Connecticut Post. This article was referenced by Benbunan-Fich (2001) in the article *Using Protocol Analysis to Evaluate the Usability of a Commercial Web Site.* While this article is published by the New York Times, and may be subject to bias, this article is included in this annotated bibliography based on the name recognition of the businesses mentioned in

the article. The audience for this annotated bibliography may be interested in the examples from these companies.

Summary. By using the redesign of International Business Machines (IBM) Corporations' e-commerce website as an example, the author explains the importance of usability testing. Prior to the redesign of their website, IBM reports that the most used feature for their website was the search tool "because people couldn't figure out how to navigate the site" (para. 2). After the redesign of the website, use of the search tool decreased by 84% and sales increased by 400%. IBM attributes the success of the website to usability design techniques such as an improved information architecture, navigation, and graphic design. Travelocity's website redesign is also used as an example. As quoted in the article, Mark Hurst, the president of Creative Good, a consulting firm that helps clients address what is known as web site usability, states that a customer's experience on the web can make or break a business. Hurst explains that over"... $3 billion was lost on the web last year because of poor design—sites not realizing that if they just made it easier for the consumer to buy, they'll make more sales" (para. 6). Sources note that website design is a crucial facet of business and that if users do not find the website easy to use, they will go elsewhere to purchase the product. To be successful, websites need to employ good usability practices.

What are the Key Categories to Include in a Website Usability Testing Model?

Literature in this section describes selected usability testing methods and the categories of website design tested by those usability testing methods. The table format is used to summarize the categories addressed in some methods.

Bachiochi, D., Berstene, M., Chouinard, E., Conlan, N., Danchak, M., Furey, T., Neligon, C., et al. (1997). Usability studies and designing navigational aids for the World Wide Web. *Computer Networks and ISDN Systems*, *29*(8-13), 1489–1496. doi:10.1016/S0169-7552(97)00027-5 http://dx.doi.org/10.1016/S0169-7552(97)00027-5

Abstract. Web designers are constantly searching for ways to improve their works. Recently published books provide such recommendations, but their quality varies greatly. This paper describes how usability testing was used to validate design recommendations. The results show a need for navigational aids that are related to the particular Website and located beneath the Browser buttons. Furthermore, usability criteria were established that limit page changes to 4 and search times to 60 seconds for information retrieval.

Credibility. At the time of publication, Daniel Bachiochi held a Bachelor degree in Computer Science and Mathematics from Boston College, in Chestnut Hill, Massachusetts. Michael C. Berstene held a Master's degree in Computer Science from Rensselaer Polytechnic Institute. Mr. Berstene managed the Human Factors Engineering group at Aetna and is an information technology professional with 25 years of experience. Elaine Chouinard held a Master of Arts degree in Experimental Psychology. Nancy Conlan is a human factors engineer and received a Master of Arts degree from Trinity College. Michael Danchak received a Bachelor of Science in Engineering from Princeton University and a PhD from Rensselaer Polytechnic Institute. Dr. Danchak, previously Dean of the School of Engineering and Sciences, at the Hartford Graduate Center was a professor at the center at the time of publication. This article is published in Computer Networks and ISDN [integrated services digital network], a peer-reviewed journal.

Summary. In this article, the authors described website navigation and some of the components that make up website navigation. Additionally, the authors describe user testing conducted at Aetna's usability laboratory and review the results of that testing. The authors note that to be successful using the website, all users, regardless of capability, must be able to navigate to the information sought (p. 1489. As a design category indentified elsewhere in this Annotated Bibliography, this article may be useful to the audience to describe some of the aspects of website navigation. The authors provide an overview of some of the current navigation aids, including back and forward buttons and hypertext links and how these aids help or hinder the user to navigate. This study looks at navigation aids provided by the website designer (those used to navigate within the website) as opposed to those provided by the web browser (those used to navigate to other websites using browser tools). Following the completion of user testing, the authors proposed a generic page layout to facilitate navigation and make several recommendations regarding website navigation tools, including the following:

- Browser tools should be augmented with a Home button, logical design structure of web pages, and navigation tools fixed at the top of the page, but beneath the browser tools.
- To be efficient in finding information, it should not take the user more than 60 seconds to find information, or make more than four page changes to find information.

Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management*, *39*(2), 151–163. doi:10.1016/S0378-7206(01)00085-4

Abstract. Despite the increasing popularity of electronic commerce, there appears to be little evidence of the methodical evaluation of the usability of commercial web sites. The usability of a web site defines how well and how easily a visitor, without formal training, can interact with the site. This paper reports the results of a research project, which applies a systematic qualitative technique known as protocol analysis or think aloud method, to examine the usability of a commercial web site. About 15 usability principles and 3 evaluation parameters (content, navigation and interactivity) were used as a framework to analyze the verbal protocols of a sample of users interacting with a greeting card web site. The protocols provided evidence of usability problems caused by crowded content, poor navigation and cumbersome interactivity. These results underscore the importance of two crucial usability goals for commercial web sites: clear path to products and transparency of the ordering process.

Credibility. Raquel Benbunan-Fich received a PhD in Management Information Systems from Rutgers University – Graduate School of Management and is currently Associate Professor of Information Systems at Baruch College. Dr. Benbunan-Fich has published over 43 journal articles (refereed), authored numerous book chapters, and presented over 32 times at academic conferences. At this time of publication, Benbunan-Fich was Assistant Professor of Information Systems at the Stillman School of Business at Seton Hall University. This article is published in a refereed journal.

Summary. This article first provides an overview of usability and methods to test usability, and includes a definition of protocol analysis, also called the talk-aloud protocol—one method of evaluating website usability. The author describes three primary criteria for evaluating websites and 15 usability principles for commercial websites (p. 154). In regards to the three evaluation criteria, *content* refers to the

presentation of information, *navigation* refers to the way users move around the application or website and move from screen to screen (for example, using buttons and frames), and *interactivity* refers to the nature of the interaction between the user and the site (p. 154). Table 4 provides a summary of those 15 principles and how they relate to three primary criteria.

Table 4

Usability Principles with Applicable Evaluation Criteria

Usability Principles	Evaluation Criteria
Simplicity of product menus	Content
Simplicity of path to products	Navigation
Supporting users with helpful categorization of merchandise	Content
Support for personalizing or narrowing product lists	Navigation
Versatility and support in the form of product pictures	Content
Support for reading and learning about products	Content
Obviousness of order buttons/links	Navigation
Supporting users by providing a compare feature	Content
Dissatisfaction with recurring security messages	Interactivity
Support for understanding the requirements of the ordering process	Interactivity
Versatility of the ordering process	Interactivity
Feedback on saving items in the shopping list	Interactivity
Obviousness and accessibility of a running total	Interactivity
Support for users who want to continue shopping	Navigation
Reversibility of actions	Navigation

Following the completion of the website usability tests described in the article, the author concludes that the usability principles and criteria listed in Table 4 are valid categories to use to "…evaluate the usability of a commercial web sites…" (p. 161). The author observes that "these factors [content, navigation, and interactivity] affect how visitors use

and perceived web-bases systems and may explain why some websites are more accepted and used than others" (p. 161).

Bolchini, D., & Garzotto, F. (2007). Quality of web usability evaluation methods: An empirical study on MiLE+. *Web Information Systems Engineering–WISE 2007 Workshops.* 481–492. doi:10.1007/978-3-540-77010 http://dx.doi.org/10.1007/978-3-540-77010

Abstract. What are the quality factors that define a "good" usability evaluation method and contribute to its acceptability and adoption in a real business context? How can we measure such factors? This paper investigates these issues and proposes to decompose the broad, general concept of "methodological quality" into more measurable, lower level attributes such as performance, efficiency, cost effectiveness, and learnability. We exemplify how to measure such attributes, reporting an empirical evaluation study of a usability inspection method for web applications called MiLE+.

Credibility. Davide Bolchini received a PhD in Communication Sciences from the University of Lugano in Switzerland and is currently an Assistant Professor (tenure track) at Indiana University. Dr. Bolchini has published over 93 peer-reviewed articles, and numerous other publications (http://mypage.iu.edu/~dbolchin/Bolchini_D_CV.pdf). Franca Garzotto received a PhD in Computer Engineering from Politecnico di Milano and is currently Associate Professor at the Department of Electronics and Information, Politecnico di Milano (Italy). She is author of over 100 international publications (http://hoc.elet.polimi.it/hoc/pages/people.php). This article is published in the proceedings of the 2007 International Conference on Web Information Systems Engineering (WISE).

Summary. This article classifies website evaluation categories by those that can be evaluated independently from the website application and those that are application

dependent. Those categories that are evaluated *independently* of the application, users, or stakeholders, can be evaluated based on technical inspections, typically using predefined guidelines or heuristics. Those categories that are application *dependent*, sometimes call user experience indicators (UEI) are evaluated in the context of use (p. 483), based on application dependent criteria including users, stakeholders, and purpose of the website. These categories are summarized in Tables 5 and Tables 6:

Table 5

Example Technical Heuristic Website Evaluation Categories

Category	**Examples of Heuristics**
Navigation	Consistency of navigation patterns
	Index backward navigation (the ability to navigate backwards without using the back arrow)
Content	Text accuracy
	Multimedia consistency
Technical performance	System reaction to user errors
	Operations management
Interface design	
Cognitive	Information overload
	Scannability
Graphics	Background contrast
	Text layout
Semiotics	Ambiguity of link labels
	Conventionality of interaction images

Table 6

Example User Experience Indicators Website Evaluation Categories

Category	**Examples of User Experience Indicators**
Content experience UEIs	Completeness
	Multilingualism

Navigation & cognitive experience UEIs	Predictability
	Memorability
Operational flow experience UEIs	Naturalness
	Recall

Chiou, W-C., Lin, C-C., & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995–2006. *Information & Management, 47*(5–6), 282–290. doi:10.1016/j.im.2010.06.00 http://dx.doi.org/10.1016/j.im.2010.06.002

Abstract. Many studies have proposed new website evaluation frameworks and criteria. We have attempted to understand and improve website evaluation through the analysis of 83 articles by classifying them into IS, marketing, and combined-approaches. Our findings showed that most early studies adopted the IS-approach but that later ones (after the burst of the dot-com bubble) shifted to a combined-approach. Our study also revealed that most papers analyzed the evaluation factors via a ranking list.

Our review showed that most studies conducted user-based surveys to examine a website, but that very few addressed strategic issues of website evaluation. We therefore proposed a strategic framework as an internal evaluation to ensure consistency between web strategy and actual website presence. The framework involved analysis of web strategy and a hybrid approach that included evaluation during three transaction phases; the framework was designed to be applied by a specific website vis-à-vis its goals and objectives through a five-stage evaluation process.

Credibility. Wen-Chih Chiou received a PhD in business management from National Sun Yat-Sen University and is an associate professor in the Department of Business Administration at the National Chin-Yi University of Technology (Taiwan). Chin-Chao Lin received a PhD in industrial engineering and enterprise information from the Tunghai

University and at the time of publication is an assistant professor at the Department of Marketing and Distribution. Chyuan Perng received a PhD in Industrial Engineering from Texas Tech University and is an associate professor in the Department of Engineering and Enterprise Information at Tunghai University.

Summary. The authors discuss website evaluations and describe the criteria, or evaluation categories, that should be tested. Recent website evaluations approaches have focused on different factors, including website usability and design, quality, user acceptance, and user satisfaction (p. 282). After a review of the literature, the authors identify the criteria, also called categories or factors, addressed in the literature pertaining to website usability testing. Through a process of content analysis, these criteria are condensed into twelve broad categories, including: (a) ease of use, (b) responsiveness, (c) personalization, (d) visual appearance, (e) playfulness, (f) responsiveness, and (g) information quality (p. 285). From the literature, the authors then identified 53 factors that were tested to evaluate these criteria. These factors were categorized into six broad categories, including: (a) place, which includes factors related to navigation; (b) playfulness, which includes factors pertaining to ease of use and convenience; (c) product, including product details, comparisons, and variety; (d) price; (e) promotion, or the organization's promotion of itself; (f) customer relations, including interactive communications, privacy policy, and customer service responsiveness (p. 287).

Davis, P., & Shipman, F. (2011). Learning usability assessment models for web sites. *Proceedings of the 16th international conference on intelligent user interfaces* (pp. 195–204). doi:10.1145/1943403.1943433 http://dx.doi.org/10.1145/1943403.1943433

Abstract. Our work explores an approach to learning types of usability concerns considered useful for the management of Web sites and to identifying usability concerns

based on these learned models. By having one or more Web site managers rate a subset of pages in a site based on a number of usability criteria, we build a model that determines what automatically measurable characteristics are correlated to issues identified. To test this approach, we collected usability assessments from twelve students pursuing advanced degrees in the area of computer-human interaction. These students were divided into two groups and given different scenarios of use of a Web site. They assessed the usability of Web pages from the site, and their data was divided into a training set, used to find models, and a prediction set, used to evaluate the relative quality of models. Results show that the learned models predicted remaining data for one scenario in more categories of usability than did the single model found under the alternate scenario. Results also show how systems may prioritize usability problems for Web site managers by probability of occurrence rather than by merely listing pages that break specific rules, as provided by some current tools.

Credibility. PhD candidate Paul Davis wrote this article as his dissertation under the guidance of Dr. Shipman, a professor at Texas A&M University. Dr. Shipman received a PhD in Computer Science from the University of Colorado at Boulder, was a 2009 ACM Distinguished Scientist, and has published at least 92 articles in various peer-reviewed journals and conference proceedings (http://www.csdl.tamu.edu/~shipman/pubs.html). This article is published in the jury-reviewed conference *Proceedings of the 16th International Conference on Intelligent User Interfaces*.

Summary. This article describes the following five categories that can be used as part of an analysis of website usability. Brief descriptions of 30 items to be check within each category are include, based on definitions obtained from http://www.usablenet.com (content accessible by subscription) (p. 199). A selected list of items include:

- Effectiveness—persons visiting the site should be able to accomplish their goals, such as finding information.
- Flexibility—there should be more than one way to reach a goal. This also means that people should be able to use the site if the page appears in browser windows of different sizes.
- Navigability—people should find it easy to learn where they are in a site as well as how to go elsewhere within the site. They should be able to remember where a page is if they return to the site.
- Satisfaction—people should not tire or become upset when using a site. They should find the experience satisfying and satisfactory.
- Efficiency—people should find that the ease of use and performance of the site satisfactory. This can include obtaining pages, determining if those pages are useful, and finding hypertext links to other pages.

Hallahan, K. (2001). Improving public relations web sites through usability research. *Public Relations Review*, *27*(2), 223–239. doi:10.1016/S0363-8111(01)00082-0

Abstract. Public relations communicators can use a variety of techniques to pretest the effectiveness of messages, including readability testing, focus groups, peer and jury reviews, experiments and field testing/test marketing. Another yet-untapped technique is usability research, a set of procedures used for more than 50 years in product and software development. Although usability is a technique that can be applied to any type of public relations message, usability research has particular relevance for enhancing the effectiveness of websites and other new interactive techniques. This article examines the nature and value of usability research, and the elements of an effective website based on

usability principles. Applications to other types of public relations communications are also discussed.

Credibility. At the time of publication, Kirk Hallahan, now a full professor (see biography at http://lamar.colostate.edu/~hallahan/), was an associate professor in the Department of Journalism and Technical Communication at Colorado State University. *Improving Public Relations Web Sites Through Usability Research* is published in the peer-reviewed Public Relations Review and has been cited at least 55 times by other authors. The references cited in this article are from journals, including the *Journal of Public Relations* and the *Public Relations Review*, reputable authors, such a Jakob Nielsen and Jared Spool, or books.

Summary. Although the phrase *usability research* is used in the title and abstract for this article, the author equates usability research to testing by using phrases such as "test[ing] military hardware" (p. 224) or testing applied to human-computer interfaces (HCIs), such as computer hardware and software (p. 224). The author defines websites as a specific type of HCI. Hallahan cites definitions of usability, such as "... how well the intended users can interact with technology to carry out an assigned activity" (p. 224), or "an interface that is workable and intuitive from the user's point of view" (p. 224), and goes on to say that "usability research strives to improve both the efficiency and effectiveness of systems" (p. 225). Hallahan notes that usability researchers have proposed the following topics of interest in website usability testing: (a) graphics do not necessarily help users, (b) text links are vital, (c) navigation and content are inseparable, (d) intentional information retrieval is different from "surfing" behaviors, and (e) people will say that they like the site even though they have trouble using it (p. 232). Additionally, Hallahan lists content and design simplicity as being *paramount* in web site

design. The following are topics related to design simplicity (p. 234): (a) system compatibility, (b) speed of use and decision making, (c) ease of navigation, and (d) accuracy of use/success of search rates.

HHS—U.S. Dept. of Health and Human Services. (2006). *Research-based web design & usability guidelines*. US Government Printing Office. Retrieved from http://www.usability.gov/guidelines/guidelines_book.pdf

Abstract. The Research-Based Web Design and Usability Guidelines (*Guidelines*) were developed by the U.S. Department of Health and Human Services (HHS), in partnership with the U.S. General Services Administration. The Guidelines were developed to assist those involved in the creation of Web sites to base their decisions on the most current and best available evidence. The *Guidelines* are particularly relevant to the design of information-oriented sites, but can be applied across a wide spectrum of Web sites.

Credibility. The Guidelines are developed and published by U.S. government departments. Contributors consisted of various experts, including some listed in this annotated bibliography, including Robert Bailey, PhD, Joseph Dumas, PhD, Melody Ivory, PhD, and Janice Redish, PhD. Each guideline is accompanied by a list of sources (references), which are listed in the Sources section at the end of the book. The sources primarily consist of literature published in peer-reviewed journals, books, or conference proceedings.

Summary. The Guidelines provide a comprehensive list of components of websites, that should be tested during website usability tests, as well as suggestions for conducting website usability tests. The authors suggest that usability specialist use the Guidelines to aid them during their review of websites and to create customized usability test check lists (p. xvii). The Guidelines are categorized by 18 different design aspects, including

hardware and software issues, page layout, navigation, links, graphics, content, and search capabilities. Each category includes topics of consideration. For example, in the category of navigation, sub-topics include navigational options, placement of menus, textual descriptions, and use of site maps. Each guideline includes an indicator of "relative importance" and "strength of evidence," based on feedback from usability experts and scholarly research (p. xix, p. xxi). The authors suggest that designers focus on those categories and sub-topics with higher rankings for relative importance and strength of evidence and then focus on categories with lower rankings, as time permit (xviii).

Krug, S. (2006*). Don't make me think: Common sense approach to web usability* (2nd ed.). Berkeley, Ca. New Riders Press.

Abstract. Five years and more than 100,000 copies after it was first published, it's hard to imagine anyone working in Web design who hasn't read Steve Krug's "instant classic" on Web usability, but people are still discovering it every day. In this second edition, Steve adds three new chapters in the same style as the original: wry and entertaining, yet loaded with insights and practical advice for novice and veteran alike. Don't be surprised if it completely changes the way you think about Web design (http://www.peachpit.com/store/product.aspx?isbn=0132809893).

Credibility. As posted on www.howto.gov (2011), a website maintained and promoted by the U.S. General Services Administration, Steve Krug has worked for 15 years as a usability expert for companies such as Apple, Bloomberg.com, Lexus.com, and NPR, to help make products and websites that people could use and enjoy. His book Don't Make Me Think has sold over 300,000 copies. Krug is the CEO of his consulting firm, Advanced Common Sense, and spends most of his time teaching usability workshops and consulting (http://www.howto.gov/training/classes/essentials-of-usability-testing). Krug

makes references to authors such as Jared Spool, Janice Redish, and Jakob Nielsen. The audience for this annotated bibliography may find this book informative and easy to read and a good background to the components of websites.

Summary. Written for the lay audience, *Don't Make Me Think*, describes usability testing, how users really use the website, categories to test, provides suggestions for working with managers not familiar with usability testing and websites, and ends with a list of recommended reading. As defined by the definition of usability testing, where "...one user at a time is shown something (whether it's a Web site, a prototype of a site, or some sketches of individual pages) and asked to either (a) figure out what it is, or (b) try to use it to do a typical task" (p. 133), usability testing includes testing the aspects of a website design. As Krug notes, "...it's never too early to start showing your design ideas to users, beginning with your first rough sketches" (p. 145). This book lists some of the aspects of website design (that should be tested), including page layout and text formatting (p. 31), information architecture and how many clicks it takes for a user to reach their destination (p. 41), content (p. 45), navigation (p. 51), and home page design and purpose (p. 103).Krug advises caution when making changes to ensure that fixing something doesn't break something else (p. 158).

Lee, Y., & Kozar, K. A. (2012). Understanding of website usability: Specifying and measuring constructs and their relationships. *Decision Support Systems, 52*(2), 450–463. doi:10.1016/j.dss.2011.10.004

Abstract. Developing a usable website is pivotal for e-business success. Researchers have devoted effort to develop metrics, guidelines and theories of website usability, yet there still is a lack of consensus on the multifaceted dimensions of website usability and lack of investigation of the nomological networks among website usability constructs.

This study first investigated the common dimensions of website usability by integrating the findings of previous studies and a focus group study with website usability experts. Instruments to measure the constructs were developed and empirically validated. Then nomological networks between website usability constructs and between those constructs and online purchase intention and purchase were examined. Three field studies including two questionnaire surveys and a causal mapping analysis were conducted. The research identified ten website usability constructs with strong psychometric properties. A number of nomological networks between usability constructs were discovered, contributing to identification of sources of variances of purchase intention and purchase behavior. Findings of this study are expected to provide useful insights for practitioners to usable websites and for researchers to better assess the effect of website usability on online customer behavior.

Credibility. Younghwa Lee is an Associate Professor of Management at the University of Northern Iowa College of Business Administration. He received a PhD from University of Colorado/Boulder in 2005. His research interest is in website usability, technology acceptance, and IT ethics and security. He is an ICIS 2003 Doctoral Consortium fellow. He has published in *Communications of the ACM, Decision Support Systems, European Journal of Information Systems, Information & Management, Journal of Management Information Systems, Journal of Organizational Computing and Electronic Commerce, MIS Quarterly* among others. Kenneth A. Kozar is a Professor of Information Systems at the University of Colorado/Boulder Leeds School of Business. His interests lie in the area of human and organizational impacts of technology. He has published in a number of journals, served two terms as an associate editor of the *MIS Quarterly*, and was the chair of the Society for Information Management's International

Paper Award Competition (http://www.sciencedirect.com/science/article/pii/S0167923611001679#bg0005).

Summary. In this article, the authors define usability and website usability and the effect of website usability on e-commerce websites. The authors explore different criteria, or factors, associated with website usability, such as content, format, search-capabilities, as proposed by different researchers. Once the different factors are identified, focus groups reviewed and categorized the factors, as shown in Appendix A of the article. The category headings were determined by the focus groups. The list provides a comprehensive view of the different categories that make up website usability and are included in website usability tests. For example, design consistency, consistency, and coherence are all grouped in the category for *consistency*. Appendix B of the article includes definitions for each of the categories. The authors note that the findings of this study are expected to be used for successfully measuring the design quality of websites and be used for building blocks for developing a strong theoretical model of website usability" (p. 459).

Lynch, P., & Horton, S. (2008). *Web style guide: Basic principles for creating web sites* (3rd ed.). New Haven, CT: Yale University Press.

Abstract. Consistently praised as the best volume on classic elements of web site design, *Web Style Guide*, now in its third edition, maintains an emphasis on fundamentals while bringing every chapter topic up-to-date.

Written for web site designers in corporations, government, nonprofit organizations, and academic institutions, the book explains established design principles and covers all aspects of web design—from planning to production to maintenance. The guide also

shows how these principles apply in web design projects whose primary concerns are information design, interface design, and efficient search and navigation.

Beginning and advanced designers will find this to be the most practical guide available.

Credibility. Patrick Horton received a Master of Science in Biology from Southern Connecticut State University. Mr. Horton as worked at Yale University in various positions, including the director of media and communication services unit for 30 years. Mr. Horton "...has authored over 100 professional papers, magazine articles, and book chapters" (http://webstyleguide.com/about-us.html). Sarah Horton has a Master of Arts in Liberal Studies from Wesleyan University. Now in private practice as a Web Consultant, Ms. Horton was the Director, Web Strategy, Design, and Infrastructure at Dartmouth College at the time of publication. She has published 54 professional papers and books. The *Web Style Guide* is in its third edition and is published by Yale University Press.

Summary. This book describes the process of website design and the categories that make up a website. Although the book does not include a specific chapter dedicated to usability testing and what to test, advice about testing is interspersed throughout the book. For example, a section describing information architecture (the content and organization of the website) advises that small prototypes of the site should be developed "...to test what it feels like to move around within the design....to test navigation and develop the user interface....[and] menus to [navigate to] content pages" (p. 29). Another section advises that use testing should be completed after the site is constructed to, in part, " ...critique the overall design and effectiveness of the site" (p. 32). One section of the book (within chapter 2) that directly discusses user testing states that, "...user testing is used throughout the design process to evaluate different design approaches by observing how well, or how poorly, they work in helping users accomplish tasks" (p. 68).

Categories of website design discussed by the authors include navigation, search capability (chap. 4), site structure and document structure (for example placement and typography used for headings) (chap. 5), page layout and order of pages (chap. 6), links (chap. 9), help and contact information (chap. 10), use of graphics (chap. 11), and the use of multimedia, for example audio, slideshows, and animation (chap. 12). For the audience of this annotated bibliography, "...this book teaches the fundamentals of interface design, information architecture, and usability without unnecessary complexity or jargon" (p. ix).

Manzari, L., & Trinidad-Christensen, J. (2006). User-centered design of a web site for library and information science students: Heuristic evaluation and usability testing. *Information technology and libraries*. Retrieved from http://faculty.mercer.edu/lewis_am/pdfs/UCD%20for%20Library.pdf

Abstract. This study describes the life cycle of a library Web site created with a user-centered design process to serve a graduate school of library and information science (LIS). Findings based on a heuristic evaluation and usability study were applied in an iterative redesign of the site to better serve the needs of this special academic library population. Recommendations for design of Web-based services for library patrons from LIS programs are discussed, as well as implications for Web sites for special libraries within larger academic library settings.

Credibility. Laura Manzari, received a Juris Doctor (JD) from Saint John's University School of Law as is currently an Associate Professor and Library Information Science Librarian at the Long Island University (C.W. Post campus) in Brookville, New York. Jeremiah Trinidad-Christensen received a Master of Science degree in Library and Information Science from Long Island University. This article was published in

Information and Technology and Libraries, a peer-reviewed journal published by the American Library Association.

Summary. In this article, the authors provide definitions of terminology related to software and website design, including human-computer interaction (HCI), usability, user-centered design (UCD), heuristic evaluation, and usability testing. The authors describe a usability study using a usability evaluation method (heuristic evaluation) and usability testing. A list of Jakob Nielsen's ten usability heuristics are included. The authors describe the process of administrating and reviewing results from usability testing. During testing, the subjects were given eight tasks to complete using the website. The tasks "...were designed to test usability of different aspects of the Web site" (p. 166). Because a library website is the subject of this study, the assigned tasks were related to assignments students might receive and information they might seek on the website. Example of a task assigned is to access a specific page on the website, determine whether the library subscribed to a specific journal, and determine whether the journal is refereed. By using the website for assigned tasks, different aspects of the website was tested. However, while the test designers did not specifically define design factors, or categories, to test, usability issues to specific design categories, for example navigation, were identified.

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence* (CHI '94), 152-158. doi:10.1145/191666.191729 http://doi.acm.org/10.1145/191666.191729

Abstract. Several published sets of usability heuristics were compared with a database of existing usability problems drawn from a variety of projects in order to determine what heuristics best explain actual usability problems. Based on a factor analysis of the

explanations as well as an analysis of the heuristics providing the broadest explanatory coverage of the problems, the following new set of nine heuristics were derived: (a) visibility of system status, (b) match between system and the real world, (c) user control and freedom, (d) consistency and standards, (e) error prevention, (f) recognition rather than recall, (g) flexibility and efficiency of use, (h) aesthetic and minimalist design, and (i) helping users recognize, diagnose, and recover from errors.

Credibility. Nielsen holds a PhD in Human-Computer Interaction from the Technical University of Denmark in Copenhagen. Nielsen has published 12 books and authored numerous articles.

Summary. In this article, the author developed a list of heuristics (a list of guidelines used when evaluating a user interface) by reviewing lists used by other researchers. Using this list, the author synthesized a new list of heuristics that are "…as good as possible at explaining the usability problems that occur in real systems" (p. 152). To develop this new list, the author collected the heuristics from other researchers in a database and then assigned a rating to each of the heuristics based on how well the heuristic explained the usability problem. The author acknowledges and accounts for the subjectivity of the ratings. Once compiled the rated the heuristics were reviewed and assigned a descriptive name. The following seven main categories were identified:

- Visibility of system status—provide feedback, indicate progress, identify cues, etc.
- Match between system and real world—speak the users, language, familiar terms, etc.
- User control and freedom—obvious way to undo, make actions reversible, etc.
- Consistency and standards—express same thing same way, uniformity, etc.
- Error prevention—design system to prevent errors, identify cues, etc.

- Recognition rather than recall—see-and-point instead of remember-and-tell, etc.
- Flexibility and efficiency of use—user-interface customizable, shortcuts, etc.
- Aesthetic and minimalist design
- Helping users, recognize, diagnose, and recover from errors

The above list can be used during a heuristic evaluation of user-interfaces "…to find their usability problems" (p. 152).

Nielsen, J. (2000). *Designing* w*eb usability: The practice of simplicity*. Indianapolis, IN: New Riders Publishing.

Abstract. *Designing Web Usability* is the definitive guide to usability from Jakob Nielsen, the world's leading authority. Over 250,000 Internet professionals around the world have turned to this landmark book, in which Nielsen shares the full weight of his wisdom and experience. From content and page design to designing for ease of navigation and users with disabilities, he delivers complete direction on how to connect with any web user, in any situation. Nielsen has arrived at a series of principles that work in support of his findings. This book is a must-have for anyone who thinks seriously about the web (http://www.peachpit.com/store/product.aspx?isbn=156205810X).

Credibility. Nielsen holds a PhD in Human-Computer Interaction from the Technical University of Denmark in Copenhagen. Nielsen has published 12 books and authored numerous articles.

Summary. In this book, Dr. Nielsen provides an explanation of what usability testing is, why websites should be tested and categories to test, by explaining good web design practices and proving examples on nearly every page. The author states that the information contained in "this book is based on observations of usability tests with about 400 users from a wide variety of backgrounds using a large number of different websites

over the last six years" (p.14). In describing the importance of website usability tests, the author states the following:

> Usability has grown dramatically in importance for web-based companies because of an inversion in the relationship between user experience and the ability to separate customers from their money. In the old world that is populated by most computer companies, customer would pay for a product first and only later take it home and discover you need a two-inch-thick book to figure out how to format chapter headings. (p. 388)

Today, "usability rules the web....if the customer can't find a product, then he or she will not buy it" (p. 9). By extension, website usability testing helps to ensure that websites are usable and that the user can readily find what they are looking for. On page 15, the author provides a high-level list of fundamental errors made when designing websites. Four major categories listed include; (a) page design, (b) content design, (c) site design, and finally (d) intranet design. Within each chapter, the author provides additional details regarding each of the four major design categories, instructions based on previous user testing, examples of evidence from user testing, as well as screen shots showing iterative examples of design changes. The book concludes with chapters on accessibility, international use, future predictions, and a conclusion.

Qi, S., Ip, C., Leung, R., & Law, R. (2010). A new framework on website evaluation. *Proceedings of the 2010 International Conference on E-Business and E-Government (ICEE),* 78-81. doi:10.1109/ICEE.2010.27 http://dx.doi.org/10.1109/ICEE.2010.27

Abstract. Information Communication Technologies (ICTs) have changed how business is conducted and consumers' purchase behavior. Online market has showed its importance to both academia and business circles. This paper analyses published website

evaluation related studies and found that the existing e-commerce literature simply does not have any commonly agreed-upon standards for evaluating overall website performance. Therefore, this research proposes a novel direction to website evaluation. A new framework has been established which suggests website assessment should be performed in three major aspects: website usefulness (technique), service quality, and physical accessibility. The expected contributions of the proposed model to academics and practitioners are discussed.

Credibility. The authors Qi, Ip, and Leung were full-time PhD students at The Hong Kong Polytechnic University in 2007. Rob Law received a PhD in Computer Science from the University of Regina and is now a professor at The Hong Kong Polytechnic University. Professor Law is an active researcher and has published at least 25 articles and books, and "...have been cited thousands of times" (http://hotelschool.shtm.polyu.edu.hk/wcms-common/temp/20120326164730088 4/RobLAW_CV.pdf). This article is published in the jury-reviewed *Proceedings of the 2010 International Conference on E-Business and E-Government*.

Summary. The article describes a website evaluation approach that focuses on website usefulness, website service quality, and website physical accessibility. These areas of focus are developed following a review of more than 100 published articles related to different aspects of website design. Under the umbrella of the three primary categories, the authors provide sub-categories that should be reviewed during website evaluations. The three primary categories in which to evaluation website usability are: (a) usefulness, (b) navigability, and (c) reliability. Is this summary, the authors note that previous studies have not always distinguished between measurements regarding technology and service

aspects and propose that this model helps to address both these aspects of website design and evaluation (p. 80).

Whitehead, C. (2006). Evaluating web page and web site usability. *Proceedings of the 44th annual Southeast regional conference.* (pp. 788–789). doi:10.1145/1185448.1185637 http://dx.doi.org 10.1145/1185448.1185637

Abstract. As the number of Web sites continue to increase, so too does the importance of Web page/Web site usability. This paper describes what constitutes Web page/Web site usability and how it can be measured.

Credibility. Christopher C. Whitehead holds a PhD in organizational management and is an assistant professor at the TSYS School of Computer Science at Columbus State University. This article is published in the proceedings of the ACM Southeast Regional Conference in 2006. The majority of references cited in this article are from peer-reviewed journals. As indicated on his website (http://csc.columbusstate.edu/whitehead/vita.asp), Dr. Whitehead has published approximately 14 papers, nine of which are related to websites or web applications.

Summary. The article *Evaluating Web Page and Web Site Usability* describes usability testing and categories that should be addressed during website usability testing. The author categorizes user testing into three categories: (a) inquiry, in which information is request from the users, such as with focus groups, interviews, questionnaires, and surveys; (b) inspection, which uses heuristic evaluation methods and cognitive walkthroughs; (c) formal, in which a formal set of tasks and goals are devised and users are watched while they perform these tasks (p. 789). The author notes that some researchers believe the user testing should be conducted using 8-12 users, but as little as five user tests may be sufficient to indicate whether the tested feature was a success or

problem (p. 789). Whitehead adds that user testing can be conducted using automated methods in addition to manual methods. The author provides a list of categories and topics that should be addressed during website usability testing, including the following:

- Learnability
- Rememberability [sic]
- efficiency of use
- Reliability
- User satisfaction
- Checkability
- Confidence
- Control
- Ease of use
- Speed
- Understanding

Conclusion

The purpose of this scholarly annotated bibliography is to identify literature that investigates usability testing for websites. Thirty-one references, consisting of articles published in peer-reviewed journals, conference proceedings, and books, are identified and reviewed. Literature is selected that addresses at least one of the following research sub-questions (Skidmore College, n.d.):

1. What is usability testing?
2. Why should websites be tested?
3. What are the key categories included in selected website usability testing methods?

As suggested by Busch et al. (2005) and Insch, Moore, and Murphy (1997), literature that meets basic criteria (as described in the Content Analysis section of this paper) is reviewed and searched for particular code words or key words, such as website usability or website testing. The results of the coding are presented in this conclusion.

As stated in the Introduction section of this paper, the primary audience for this annotated bibliography is managers with influence over the design and content of websites and website applications. Secondary audiences include newcomers to the field of website design and current website designers interested in learning more about website usability testing and its affect on website design, in order to help them deliver more usable websites (Tullis & Albert, 2008, Preface).

What is Usability

To understand website usability testing, it is important to understand what usability is and how usability testing relates to usability. In the literature included in this annotated bibliography, the authors describe usability using similar terms. For example, Hallahan (2001) refers to usability as "...how well the intended users can interact with a technology to carry out an assigned

task [where the focus of performance is on the technology, not the user]" (p. 224), and Levi (2008) defines usability as "...the degree to which a given piece of software assists the person sitting at the keyboard to accomplish a task, as opposed to becoming an impediment..." (para. 3). Christiansen and Frøkjaer (2010) relate usability to effectiveness, efficiency, and satisfaction (p. 119). As reported by Downing and Liu (2011), usability is defined by the International Organization of Standardization (ISO) as "the extent to which a product or a service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of user" (p. 144). While the ISO definition applies to any product or service, Whitehead (2006) notes that the definition applies equally well to websites (p. 788).

Writing about website usability, Nielsen (2003) describes usability as a "...quality attribute that assesses how easy user interfaces are to use" (What section, para. 1), and provides the following five principles that define usability:

- **Learnability**: How easy is it for users to accomplish basic tasks the first time they encounter the design?
- **Efficiency**: Once users have learned the design, how quickly can they perform tasks?
- **Memorability**: When users return to the design after a period of not using it, how easily can they reestablish proficiency?
- **Errors**: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- **Satisfaction**: How pleasant is it to use the design? (Nielsen, 2003, What section, para. 2)

Whitehead (2006) recounts McLaughin and Skinner's principles for determining the usability of a website:

- **Checkability**: The system has or allows checks that ensure the correct information is

going in and going out of it.

- **Confidence**: Users have confidence both in their capability to use the system and in the system itself.
- **Control**: Users have control over the operation of the system, particularly of the information fed into and out of the system.
- **Ease of Use**: The system is easy to use.
- **Speed**: The system can be used quickly.
- **Understanding**: The system and its outputs are understandable. (p. 788)

Usability may be summarized as a service or product's design attributes that make it easy for a user to use the service or product easily, effectively, and with satisfaction (Levi & Conrad, 2008, para. 3). Additionally, there are various metrics by which usability may be measured (Whitehead, 2006, p. 788). Or, "stated another way, usability focuses upon whether the system can be used to achieve some desired goal" (Hallahan, 2001, p. 225).

What is Usability Testing

With an understanding that usability refers to the design attributes that affect the user's ability to use a product or service, one might next ask how and why websites and other products and services are tested for usability. This section discusses usability testing and how it is performed. The rationale for performing usability testing is described later in this document.

Terminology describing usability testing is used in different ways by researchers, resulting in potential confusion for readers. For example, Hallahan (2001) refers to usability testing as *usability research* (p. 224) and Nielsen (2000) refers to usability testing as *evaluation*. Palmquist (2001) observes that because the terms (a) *usability engineering* (which involves the user during the entire design process), (b) *usability analysis* or *inspection* (which involves asking

a user to perform a simple task on a prototype or nearly completed system), and (c) *usability testing* (assessing the completed, or nearly completed, product in a realistic setting) are frequently used when describing usability, the distinction between these terms often becomes blurred (p. 126). Following the convention used by Levi and Conrad (2008), the term usability testing is used is this study to broadly "...describe all methods of assessing or measuring system usability, regardless of participant population" (para. 13).

The purpose of usability testing. As described by Levi and Conrad (2008), "usability testing is the process by which human-computer interaction characteristics of a system are measured and weaknesses are identified for correction" (p. 7). "One of the main purposes of usability testing is to identify issues that keep users from meeting the usability goals of the Web site [or, by inference, a product or service]" (HHS, 2006, p. 203), and "...to cause beneficial improvements to the user interface..." (Redish, Bias, Bailey, Molich, Dumas, & Spool, 2002, p. 887). Similarly, Manzani and Trinidad-Christensen (2006) describe usability testing as "...an empirical method for improving design" (p. 165). A secondary purpose is to provide the design team with information to assist with "...eliminating unnecessary distractions and reduce errors before the product reaches the final development and implementation stages" (Hallahan, 2001, p. 225).

Terms to describe usability testing methods. Authors included in this study use a variety terms to describe usability testing methods, such as qualitative, quantitative, summative, formative, formal, discount, comparative, and informal. Nielsen and Molich (1990) use the following terms to describe four approaches to test a website user interface:

- Formally, by some analysis technique using exact models and formulas to calculate usability measures
- Automatically, by a computerized procedure

- Empirically, by experiments with real users
- Heuristically, by simply looking at the interface and passing judgment according to one's own opinion (p. 249)

Redish et al. (2002) and HHS (2006) use a slightly different set of four terms to describe usability testing methods:

- Automated, by a computerized procedure
- Inspection, such as a heuristic evaluation or cognitive walk-through
- Performance, where representational users complete tasks
- Operational, where information about the system, such as download speeds is gathered

Whitehead (2006) uses three terms to describe usability testing methods:

- Inquiry, characterized by requesting information from the user (surveys, focus groups, interviews etc.)
- Inspection, such as a heuristic evaluation or cognitive walk-through
- Formal, where users attempt to complete a list of formal goals and/or tasks (p. 789)

Hallahan (2001), discussing website usability testing specifically describes usability testing methods under two broad terms, with sub-terms in one:

- Laboratory testing conducted with users who are tested in the laboratory setting
- Alternative assessment techniques conducted outside the laboratory setting (p. 225)
    - Inspection (experts assess the website with methods including heuristic evaluations and cognitive walkthroughs)
    - Inquiry-based
    - Other, such as card sorting, website analysis, and reputation management software (p. 228)

While the majority of the terms presented above to describe usability testing methods are applicable in a variety of industries, the nature of the web lends itself to additional terms including (a) remote testing, (b) mining logs for website use analysis, and (c) employing reputation management software (special software designed to collect feedback from users and allow users to submit comments, either publically on the website or privately to the organization) (Levi & Conrad, 2008, p. 43; Whitehead, 2006, p. 228).

Usability Testing Methods and Related Techniques

Lynch and Horton (2008) define usability testing as "...a controlled and directed observation of user behaviors when working with a design" (p. 68). While traditional usability testing conducted in a laboratory is typically performed by someone not associated with the design team, depending on the method of testing, "usability testing can be performed with developers, human-computer interface experts, or representative end users" (para. 13).

The following describes several usability testing methods and related techniques in greater detail. Unless noted, testing techniques can be used for usability testing in general and for website usability testing specifically. These techniques "...can range from the rigorously structured to highly informal, from quite extensive to virtually fee, and from time-consuming to quick" (Levi & Conrad, 2008, para. 7).

Laboratory usability testing. Described in the literature as an empirical usability testing method (HHS, 2006, p. 203; Levi & Conrad, 2008, para. 13; Nielsen & Molich, 1990, p. 249; Redish et al., 2002, p. 886), laboratory testing with users may be one of "...the classic, most sophisticated, and most reliable approach[s] to usability assessment..." (p. 227). Krug's (2006) description of laboratory usability testing follows:

> Usability testing has been around for a long time, and the basic idea is pretty simple: If you want to know whether your software or your Web site or your VCR

> remote control is easy enough to use, watch some people while they try to use it and note where they run into trouble. Then fix it, and test it again.
>
> In the beginning, though, usability testing was a very expensive proposition. You had to have a usability lab with an observation room behind a one-way mirror, and at least two video cameras so you could record the users' reactions and the thing they were using. You had to recruit a lot of people so you could get results that were statistically significant. It was science. It cost $20,000 to $50,000 a shot. If didn't happen very often. (pp. 135-136)

When testing software applications using the laboratory method, including websites, special screen capture software is typically employed to record keystrokes and mouse movements (Hallahan, 2001, pp. 225-226; HHS, 2006, p. 196).

Hallahan (2001) points out that "most usability tests rely upon triangulation, that is combining several data gathering techniques—including quantitative and qualitative measures" (p. 226). He explains that researchers complete their own scoring sheets while observing the testing in the laboratory. This data is then combined with the videotaped recordings, comments from the participant, questionnaires, and keystroke and mouse-click data. "Together these multiple measures provide richer insights into the user's response than would be possible with a single measure" (Hallahan, 2001, p. 226).

Some of the problems associated with laboratory testing include: (a) it is expensive to set up a full usability laboratory, (b) mistakes can be made in planning and conducting the tests, (c) problems with the validity and reliability of measures, (d) representative users were not identified, and (e) test results are misinterpreted (Hallahan, 2001, p. 227). Table 7 provides a summary of the procedures for conducting usability tests in a laboratory:

Table 7

Procedures for Conducting Usability Tests in a Laboratory

1. Develop a research question
2. Conduct a task analysis—identify what user-participants will do in the test
3. Select and plan measurements—write a scenario, arrange and become familiar with equipment, prepare the interview questionnaire, prepare check sheets to speed recordings of observations. Also obtain human subjects research approval, if required.
4. Identify and recruit subjects
5. Collect data—based on observation data, recorded keystrokes and mouse movements, transcripts of verbalizations, end-of-test questionnaire responses.
6. Analyze and interpret the data—including subjective analysis by observers, content analysis of videotape and audio tape transcripts, results of questionnaire.
7. Summarize results and make conclusions.
Adapted from: Donald E. Zimmerman and Michel Lynn Muraski, "Usability Testing—An Evaluation Technique," in The Elements of Information Gathering. A Guide for Technical Communicators, Scientists and Engineers (Phoenix, Ariz: Oryx Press, 1995), pp. 180-187. (Hallahan, 2001, p. 227)

While laboratory usability testing is known for being expensive to conduct, Krug (2006) and Nielson and Molich (1990) believe that adjustments can be made to lower the cost. For example, Nielsen and Molich (1990) proposed a *discount evaluation* method (described under Heuristic Testing in this paper) and Krug (2006) describes what he calls the *lost-our-lease* method of testing usability. In this method, the laboratory setting is improvised with equipment from home, fewer participants are required and are paid a low stipend for the time, and the development team debriefs over lunch the same day (as opposed to waiting for a lengthy report and analysis as typical for traditional laboratory testing) (Krug, 2006, p. 137). Table 8 summarizes the differences between traditional laboratory testing and lost-our-lease testing, as noted by Krug (2006):

Table 8

Differences between Traditional and Lost-Our-Lease Testing

	Traditional Testing	**Lost-Our-Lease Testing**
Number Of Users Per Test	Usually eight or more to justify the set-up costs	Three or four
Recruiting Effort	Select carefully to match target audience	Grab some people. Almost anybody who uses the Web will do.
Where To Test	A usability lab, with an observation room and a one-way mirror	Any office or conference room
Who Does The Testing	An experienced usability professional	Any reasonably patient human being
Advance Planning	Tests have to be scheduled weeks in advance to reserve a usability lab and allow time for recruiting	Tests can be done almost any time, with little advance scheduling
Preparation	Draft, discuss, and revise a test protocol	Decide what you're going to show
What/When Do You Test?	Unless you have a huge budget, put all your eggs in one basket and test once when the site is nearly complete	Run small tests continually throughout the development process
Cost	$5,000 to $15,000 (or more)	About $300 (a $50 to $100 stipend for each user and $20 for three hours of videotape)
What Happens Afterwards	A 20-page written report appears a week later, then the development team meets to decide what changes to make	Each observer writes one page of notes the day of the test. The development team can debrief the same day.

In summary, usability testing conducted in the laboratory potentially provides the most comprehensive feed back to researchers. However, the traditional version of conducting testing is expensive and time consuming and may not provide feedback to the development team in a timely manner. As an alternative, modifications to formal laboratory testing, such as those

proposed by Krug (2006), may provide useful results in a more timely manner, allowing the organization to perform more frequent testing throughout the design process. While Krug (2006) recommends that the preferred method of conducting usability testing is by a usability professional, if the organization cannot afford a professional, he believes that these lost-our-lease techniques are better than no testing at all (p. 137).

Heuristic usability testing. Hallahan (2001) categorizes usability testing methods into two broad categories—(a) laboratory testing (also categorized as empirical testing by Nielsen and Molich (1990)) and (b) alternatives to laboratory testing, including inspection methods and inquiry-based methods. This section addresses heuristic evaluation, one of the inspection methods of usability testing.

Heuristic evaluation is an informal inspection method for finding certain types of usability problems in the user interface design (HHS, 2006, p. 2006). Using this method, small groups of evaluators examine the interface for design problems and judge its compliance with recognized usability principles (Hallahan, 2001, p. 228; HHS, 2006, p. 200; Hollingsed & Novick, 2007, p. 250; Nielsen & Molich, 1990, p. 249; Manzari & Trinidad-Christensen, 2006, p. 164; Redish et al., 2002, p. 886). While early collections of usability guidelines contained over one thousand rules to follow, Nielson and Molich (1990) present a much smaller list of nine principles (p. 249). Hallahan (2001) observes that "Nielsen stresses the importance of focusing heuristic evaluation on key criteria, rather than a litany of every possible problem" (p. 228). The nine heuristic techniques provided by Nielsen and Molich (1990) are as follows:

- Use simple and natural dialogue
- Speak the user's language
- Minimize memory load
- Be consistent

- Provide feedback
- Provide clearly marked exits
- Provide shortcuts
- Provide good error messages
- Prevent errors in the first place (p. 249)

After completing a series of experiments pertaining to heuristic evaluations, Nielsen and Molich (1990) conclude the heuristic evaluation is a valid approach for performing usability evaluations, but it is also difficult to do well. Therefore, approximately five evaluators should perform the evaluation, but more evaluators than this do not provide significantly better results (p. 255). However, Redish et al. (2002) propose that twelve evaluators provide better results (p. 886). Some of the advantages of heuristic evaluation are: (a) it is relatively inexpensive, (b) it is intuitive to learn, and (c) it can be performed quickly without advance planning, and (d) it can be used early in development (p. 255). Disadvantages include that (a) the results may be biased by the evaluator's beliefs, (b) issues that are not related to usability problems may be identified, and (c) results may not lead to breakthroughs in the design process (p. 228).

In summary, heuristic evaluations can be performed quickly at any time in the design process for anyone following a set of guidelines. However, because evaluators tend to identify different issues, several evaluations should be performed by five to twelve people. Although easy to perform, heuristic evaluations may identify issues that are not directly related to usability, and results may be biased based on the evaluator's beliefs or mindset.

Inquiry-based usability testing. One of the alternatives to laboratory usability testing, as categorized by Hallahan (2001), is inquiry-based testing methods. "Inquiry-based research involves directly observing and asking questions of users to identify problems and assess

effectiveness outside of the laboratory" (p. 228). Some of the techniques used for inquiry-based testing include:

- Structured field interviews
- Ethnographic observations
- Interviews
- Focus groups
- Surveys
- Questionnaires (Hallahan, 2001, p. 228)

One disadvantages in the inquiry-based approach is that the participant may not have design training, and "...can mislead usability researchers and distract them from quickly identifying workable solutions to defects..." (Hallahan, 2001, p. 230).

Why Test Usability

This section of this paper discusses the rational for conducting usability testing in general and website usability testing specifically. At the heart of usability testing is profitability. As Tullis and Albert (2008) observe, "championing usability in a business setting is often geared toward increasing revenues and/or decreasing costs" (Section 1.3, para. 5). If usability testing did not impact profitability, why else would computer firms "...spend millions of dollars testing software and hardware usability ..." (Hallahan, 2001, p. 255)? Table 9 provides an overview of topics described in the literature to address the rationale for performing usability testing.

Table 9

Rationale for Performing Usability Testing

Rationale	Citations
Design considerations (navigation, download speed, and	Benbunan-Fich (2001); Black (2002);

usability practices)	Downing & Liu (2011); HHS—U.S. Dept. of Health and Human Services (2006); Krug (2006); Levi & Conrad (2008); Spool (2009); Tedeschi (1999)
Can't find it/Will leave	Krug (2006); Nielsen (2000); Tedeschi 1999)
Profit/Purchases and sales	Benbunan-Fich (2001); Black (2002); Hallahan (2001); Krug (2006); Nielsen. (2000); Spool (2009); Tedeschi (1999)

Rationale for usability testing in general. As noted by Hallahan (2001), "Usability experts suggest a variety of reasons for conducting usability tests. These include:

- The assurance of ease [of use] and satisfaction by users
- Establishment of benchmarks for future versions
- Minimization of service, training, and support costs
- Increased use and support of the product (or site) by users
- Imperative to be competitive" (p. 230)

Levi and Conrad (2008) explain that the website development lifecycle is the same as that used for traditional software , including requirement gathering, analysis, design, implementation, testing and deployment, "and that just as traditional software development should have a functionality and usability component, so should Web development efforts" (para. 2). Historically, human-computer interaction (HCI) principles developed over the past 15 years have focused "...on end users and their tasks, empirical measurements of system usage, and iterative development" (Levi & Conrad, 2008, para. 5). "Usability testing is the process by which the

human-computer interaction characteristics of a system are measured, and weaknesses are identified for correction" (Levi & Conrad, 2008, para. 7).

Rationale for testing websites. The difference between traditional software and website applications is that if the user does not like the site, for example, they cannot find what they are looking for or they cannot figure out how to use a feature, they will just leave (Bachiochi, Berstene, Chouinard, Conlan, Danchak, Furey, Neligon, et al., 1997, p. 1491; Black, 2002, p. 9; Levi & Conrad, 2008, para. 9; Nielsen, 2000, p. 10). As Black (2002) states, "the web has...brought technology to your grandma....You can't train her. And if she can't figure it out, she'll just go away" (para. 9).

As a specific types of software (Hallahan, 2001, p. 224), websites have commonalities and differences from other software systems, which must be taken into account when performing usability testing (Levi & Conrad, 2008, para. 54). Levi and Conrad (2008) provide the following list of unique elements to consider for website development:

- *A highly diverse user population which is non-trivial to predict or measure*. This makes finding a "representative" set of test participants difficult.
- *A highly diverse set of end-user computer configurations, including hardware, systems software, and browsers*. Ideally, usability testing will be performed from multiple client machines using multiple browsers. In practice, this geometrically increases the number of required test machines and test participants, and is usually not feasible.
- *A wide disparity in connectivity speed and bandwidth*. Again, in an ideal world testers would have enough client hardware and test participants available to cover the possible permutations. Again, this is typically not feasible.
- *A deployment environment which gives the illusion of being much more powerful than it actually is*. Since most browsers run in a windowed environment, and most Web pages

include graphics, different size fonts, etc., the inexperienced user is misled into expecting the full functionality of a graphical user interface application. Java and ActiveX applets may move the capabilities of a Web sites closer to such expectations, but testers must expect and account for user disappointment.

- *A deployment environment that blurs the distinction between the site content and the browser used to access this content*. Test participants frequently comment on deficiencies in the browser being used, and may not understand the distinction between browser and Web site. Though this may be helpful in developing an understanding of users dissatisfactions, the site designer typically has no control over browser development. (para. 54-57)

As Chiou, Lin, and Perng (2010) note, website visitors are not just users. They are potential clients and customers (p. 286), but, unfortunately, not all websites are successful at turning users into customers (Chiou, Lin, & Perng, 2010, p. 282). Downing and Liu (2011) state that more and more companies are going online to conduct business and ecommerce and ecommerce will continue to grow (p. 144). For many businesses, losing customers because of poor design could be catastrophic (Levi & Conrad, 2008, para. 9) and indeed, as noted by Black (2002), currently, ecommerce sites lose nearly half of their potential sales because visitors can't figure out how to use them (para. 10). As Downing and Liu (2011) note, research indicates that online purchase behavior is directly related to website usage (p. 145), and is a significant antecedent of purchase behavior (p. 144). Bottom line, according to Benbunan-Fich (2001), "...better usability will result in more efficient interaction between the user and the site and will increase the probability that the user will return and/or make a purchase" (p. 151). Businesses and organizations, then, cannot afford to not ensure that their websites are easy to use, efficient, and satisfying to users (and customers).

Spool (2009) relays the story of the $300 million dollar button. By performing usability testing, a problem, or rather part of the website design, was identified that users found off putting–off putting enough that they did not stick around to complete purchases, even once items were in the electronic shopping cart (having the item in the shopping cart is important because it signifies that the users most likely were interested in making a purchase). By fixing one small button, the total number of customers completing purchases went up 45% and those sales resulted in an additional $15 million the first month. Later analysis revealed that the customers who did make purchases prior to the fix, had many multiples of user names and password stored on the system—data that cost the business to store and support. Fixing this one button also solved that problem, which resulted in reduced costs, thus also affecting the bottom line. Usability testing is not just a good idea—it could result in million dollars of sales.

Categories in Selected Usability Testing Methods

This section presents categories described in selected usability testing methods. Table 10 summaries ten usability testing categories addressed across the different methods selected for review in this annotated bibliography. The ten categories include:

- Content relevance
- Ease of use and decision making
- Graphics
- Interactivity between the site and the user
- Interface design
- Links
- Navigation/Menus
- Page layout/Text formatting

- Search tools
- Technical performance

Category schemes presented by Chiou, Lin, and Perng (2010), Lee and Kozar (2012), and Nielsen (1994) are presented in separate tables, as a way to include more detail. Although the terms used to describe the methods vary (for example, categories, dimensions, criteria), the meanings are similar. As indicated in Table 10, the categories most frequently listed address (a) navigation, (b) search features, and (c) content.

Table 10

Usability Testing Categories Listed by Citation

		Categories									
Citation	**Usability Testing Method**	**Content Relevance**	**Ease of Use and Decision Making**	**Graphics**	**Interactivity (between site & user)**	**Interface Design**	**Links**	**Navigation/ Menus**	**Page Layout/Text Formatting**	**Search Tools**	**Technical Performance**
Bachiochi, Berstene, Chouinard, Conlan, Danchak, Furey, Neligon, et al. (1997)	Non-specific							X			
Benbunam-Fich (2001)	Inspection (Protocol Analysis)	X			X			X			
Bolchini & Garzotto (2007)	Inspection (Heuristic)	X				X		X			X
Chiou, Lin, & Perng (2010)	See Table 13 and Table 14										

		Categories									
Citation	**Usability Testing Method**	**Content Relevance**	**Ease of Use and Decision Making**	**Graphics**	**Interactivity (between site & user)**	**Interface Design**	**Links**	**Navigation/ Menus**	**Page Layout/Text Formatting**	**Search Tools**	**Technical Performance**
Hallahan, (2001)	Inspection (Heuristic)		X							X	X
HHS—U.S. Dept. of Health and Human Services (2006)	Guidelines	X		X		X	X	X	X	X	X
Krug (2006)	Non-specific	X			X			X	X		
Lee, & Kozar (2012)	See Table 12										
Lynch & Horton (2008)	Non-specific		X	X	X	X	X	X	X	X	
Manzari & Trinidad-Christensen (2006)	Non-specific	X					X	X	X	X	
Nielsen (1994)	Inspection (Heuristic)	See Table 11									

		Categories									
Citation	**Usability Testing Method**	**Content Relevance**	**Ease of Use and Decision Making**	**Graphics**	**Interactivity (between site & user)**	**Interface Design**	**Links**	**Navigation/ Menus**	**Page Layout/Text Formatting**	**Search Tools**	**Technical Performance**
Nielsen (2000)	Non-specific	X							X		
Qi, Ip, Leung, & Law (2010)	Non-specific	X	X			X		X	X		X

Table 11 provides Nielsen's (1994) list of seven categories to use for heuristic evaluations.

Table 11

Nielsen's (1994) Seven Main Categories for Use in Heuristic Evaluations

	Category
1	Visibility of system status—provide feedback, indicate progress, identify cues, etc.
2	Match between system and real world—speak the users, language, familiar terms, etc.
3	User control and freedom—obvious way to undo, make actions reversible, etc.
4	Consistency and standards—express same thing same way, uniformity, etc.
5	Error prevention—design system to prevent errors, identify cues, etc.
6	Recognition rather than recall—see-and-point instead of remember-and-tell, etc.
7	Flexibility and efficiency of use—user-interface customizable, shortcuts, etc.
8	Aesthetic and minimalist design
9	Helping users, recognize, diagnose, and recover from errors

Table 12 provides the results of Lee and Kozar's (2012) study in which they conducted a review of the literature with the purpose of identifying various website usability constructs (for example, navigation, consistency, etc.). Lee and Kozar (2012) include a list of related studies as well as question used to measure the construct. For example, for the construct *learnability*, one of the questions used to measure learnability is *I can remember how to reach the same page I visited next time*.

The author finds Lee and Kozar's (2012) usability constructs to be more helpful than Nielsen's (1994), because they describe the construct, or the what, and identify specific functionality to review. if a novice reviewer used Table 11 as a guide for reviewing a website, they may not know where to look or how to determine whether error prevention, for example, was used or not.

Table 12

Lee and Kozar (2012) Constructs, Definitions, Related Studies, and Measurement Questions

Construct	Definition	Related Studies	Usability Measurement Question
Consistency	Consistent location of page components within and across pages	Empirical studies have found that highly consistent websites (or interfaces) decrease error rates and learning time, and increase performance and user satisfaction. Shneiderman mentioned that striving for consistency is the first golden rule of interface design. Lohse and Spiller asserted that consistent menu bars and links significantly influence online consumer behavior.	(1) The website repeats the same structure, components, and overall look across pages. (2) The website contains similar components across web pages. (3) Web pages in the website are consistently designed. (4) Each web page on the website is of similar design. (5) The website adheres to rules and standards of other online shopping sites.
Navigability	Capability to provide alternative interaction and navigating techniques	Navigation is an important design element, allowing users to acquire more of the information they are seeking and making the information easier to find. Diverse features of a search engine and multiple navigation options are examples of navigability that can implemented. Researchers indicated that hard to navigate websites with restricted flexibility are a major usability problem.	(1) The website provides multiple search features (e.g., search engine, menu bar, go-back-and-forward button, etc.) to obtain the target information. (2) The web page that I am looking for can be reached through multiple pathways. (3) There are multiple ways to access the web page that I am looking for and/or return to shopping menus. (4) It is very easy to locate what is needed in this website.

Construct	Definition	Related Studies	Usability Measurement Question
			(5) The website keeps the user oriented as they shop.
Supportability	Additional information and support mechanisms readily available to enhance the website use experience	Consumers want to receive careful, continuous, and convenient service during navigation. Websites with just-in-time online support (e.g., FAQs, 24x7 online service), and online discussion boards were found to have more consumer visits and sales identified consumer support as a major construct for success in online shopping	(1) While visiting the website, I feel that I can get just-in-time support anytime I need it. (2) The website provides features to ask for help anytime I need. (3) Getting support through a series of options is easy and convenient.
Learnability	Easy to learn the main functionality and gain proficiency to complete the tasks	Online consumers want to spend less effort when they revisit websites and perform similar tasks. Liu and Arnett mentioned that increasing consumers' abilities to learn how to browse and find relevant information are directly related to their online satisfaction	(1) The contents provided by the website are easily understood. (2) The website is designed for easy understanding. (3) I can easily remember how to reach the same page when I visit next time. (4) As time passes, I am more accustomed to the website with less effort.
Simplicity	Provision of minimum contents and functions within a website	Simplicity in website design promotes fast system response, low cognitive load to navigate, low error rates, and less disorientation. Companies have been misled to develop a website with functions (e.g., multimedia) and copious content, but it was found that the complexity of most current websites causes negative effects on online users. Rosen and Purinton pointed out that simplicity of website design not only makes	(1) The structure of the website is succinct. (2) I can comprehend most components of a page within seconds. (3) The website has components that are not necessary. (4) There are redundant components in the website.

Construct	Definition	Related Studies	Usability Measurement Question
		the site more appealing, but also makes it far faster to load	
Interactivity	Website's ability to create vivid interaction and communication with users	Previous research has found significant effects of interactivity on user performance. For instance, Ariely pointed out that an interactive website provides more knowledge, comfort, satisfaction, and trust. Palmer indicated that interactivity is a key capability for a good website design, and Schubert and Selz suggested interactivity as a crucial component of web assessment.	(1) The website provides an appropriate amount of interactive features (e.g., graphics, pop-up windows, animation, music, voices). (2) The website contains components to help the interaction between it and consumers. (3) Interactive features of the website are vivid and evoke responses. (4) The website provides features for interactive communication between consumers, or between consumers and the online company.
Readability	Extent to which website components are well organized and easy to read and understand	Readability is affected by size of displayed characters, the relative contrast between text characters and page background, the sets of colors as well as the structural composition of the websites. Nielsen suggested using 50% less text than on paper since it is more difficult to read from a website than on paper.	(1) The website's wording is clear and easy to understand. (2) The website has enough white space (or margins) to make it readable. (3) Every page contains the appropriate amount of components to fit into a page. (4) The website uses colors and structures that are easy on the eyes.
Content relevance	Extent to which the content is up-to-date and pertinent	Content relevance is directly associated with the appropriate range and scope of content and content newness. Studies have indicated the importance of content	(1) The website contains in-depth information. (2) The website provides up-to-date

Construct	Definition	Related Studies	Usability Measurement Question
		relevance to online shopping. For example, Cole et al. suggested that the completeness and relevance of substantive information on a website is assessed by depth, breadth, and newness of its content.	information. (3) The scope of information provided by the website is appropriate. (4) The information provided by the website is accurate.
Credibility	A holistic concept that covers an online user's perception of security, privacy, and reliability during the navigation	Previous studies indicated that credibility is one of the biggest obstacles to e-commerce. For example, Nielsen pointed out that online consumers do not disclose their personal and financial information until they are convinced that the website is secure. Nielsen also insisted that websites should implement multiple features (e.g., encryption, privacy seals) to assure security and privacy of online shopping. Reliability is a part of credibility. Unstable systems frustrate consumers and diminish the consumption experience.	(1) I feel safe in my transactions with the website. (2) I trust the website to keep my personal information safe. (3) I trust the website administrators will not misuse my personal information. (4) The website is stable to use. (5) Services are routinely delivered as promised. (6) The website provides detailed information about security features.
Telepresence	Sense of presence in a virtual environment created by a computer/communication medium	Telepresence theory assumes that media can provide a feeling of presence by approximating real-world feelings, which consumers can use or may consider a standard for assessing their online experience. Studies pointed out that online consumers want to feel and touch the products, and communicate with sellers similar to what they experience in physical markets.	(1) I felt empathy with the website. (2) I feel I have personal ties to the website. (3) I feel as though I am emotionally connected to the website. (4) I feel as though I am taking part with the website.

Tables 13 and 14 present the results of Chiou, Lin, and Perng's (2010) review of the literature to identify categories to test during website usability testing. Over 83 journal articles were reviewed to identify 53 criteria for website usability testing (see Table 14); these criteria are consolidated into twelve categories (see Table 13).

Table 13

Chiou et al. (2010) 12 Categories of Website Usability Testing

Category	Description
Ease of use	Usability, accessibility, navigability, and logical structure
Responsiveness	Accessibility of service, e-mail service, reply to customer, contact information, and intuitive online help
Fulfillment	Order process, accuracy of service promise, billing accuracy, online booking process, and confirmation, on-time delivery
Security/Privacy	Information protection, online purchase security, and privacy statement
Personalization	Individualized attention, customization of offerings and information
Visual appearance	Attract attention, convey image, and aesthetics
Information quality	Variety, scope, currency, conciseness, accuracy, authority, reliability, and uniqueness
Trust	Brand recognition, consistency, intentions, and credibility
Interactivity	Interactive features and communication (FAQs, guest books, chat)
Advertising/persuasion	Marketing, promotional content, suggested products, recommendation, and incentives
Playfulness	Enjoyment, fun, pleasure, and flow
Technology integration	New technology and integration

Table 14

Chiou et al. (2010) 53 Criteria for Website Usability

Factors/Criteria	**Factors/Criteria**
Place	Product
Ease of navigation	Product details
Content relevancy and usefulness	Product comparison
Appealing and consistent style	Product search or assortment
Security protection	Product variety
Logical structure	Hierarchical product category
Ease of online transaction	Product quality
User-friendly interface	Price
Comprehensive content coverage	Price details
Loading and processing speed	Competitive price
Up-to-date content	All relevant charges details
Proper multimedia	Price comparison
Well and quick linkage	Promotion
Searching mechanism	Promotion campaign
Ease of access	Reputation and credibility of the site
Easy to understand and read	Company and brand recognition
Reliable and innovative system	Purchasing guarantee
Accuracy	Advertising and banner
Easy to find target information	Customer relations
Online assistance and help	Interactive communications
Data retrieve mechanism	Customized service
Playfulness	Privacy policy
Convenient payment methods	Quick response to customer
Know the present location	Customer service support
Overview of selected items	Member community
Easy to cancel or modify order	Order status inquiry and tracking
	Valuable bundles or product suggestion
	Delivery product as promised
	Customized offerings
	Convenient delivery options
	Ease of registration
	Easy to return product

References

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal*, *11*(4), 325–338.

Allen, M., Currie, L. M., Bakken, S., Patel, V. L., & Cimino, J. J. (2006). Heuristic evaluation of paper-based web pages: A simplified inspection usability methodology. *Journal of Biomedical Informatics*, *39*(4), 412–423. doi:10.1016/j.jbi.2005.10.004

Al Neimat, T. (Oct 24, 2005). Why IT projects fail. Project Perfect. http://www.projectperfect.com.au/info_it_projects_fail.php

American Heritage Dictionary of the English Language. (2006). 4th ed. Houghton Mifflin Company, Boston, MA.

APA Style Guide. (2010). *Publication manual of the American psychological association* (6th ed.). Washington, DC.

Atomic Design. (n.d.). Website usability testing. Retrieved on May 10, 2012, from http://www.atomicdesign.net/usability-testing.aspx

Bachiochi, D., Berstene, M., Chouinard, E., Conlan, N., Danchak, M., Furey, T., Neligon, C., et al. (1997). Usability studies and designing navigational aids for the World Wide Web. *Computer Networks and* I*SDN Systems*, *29*(8-13), 1489–1496. doi:10.1016/S0169-7552(97)00027-5 http://dx.doi.org/10.1016/S0169-7552(97)00027-5

Baxter, J. (2009). Content analysis. In R. Kitchin & N. Thrift (Eds.) *International Encyclopedia of Human Geography* (pp. 275-280). doi:10.1016/B978-008044910-4.00415-6 http://dx.doi.org/10.1016/B978-008044910-4.00415-6

Bell, C., & Smith, T. (2009). *UO Libraries: Critical evaluation of information sources*. University of Oregon Libraries. Retrieved from http://libweb.uoregon.edu/guides/findarticles/credibility.html

Benbunan-Fich**,** R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management, 39*(2), 151–163. doi:10.1016/S03787206(01)00085-4 http://dx.doi.org/10.1016/S0378-7206(01)00085-4

Bevan, N., Kirakowski, J., & Maissel, J. (1991). What is usability. *Proceedings of the 4th International Conference on HCI*. Retrieved from http://www.usabilitynet.org/papers/whatis92.pdf

Black, J. (2002). Usability is next to profitability. December 4, 2002, special report. *Bloomberg Business Week Online*. New York, NY. Retrieved on May 27, 2012, from http://www.businessweek.com/technology/content/dec2002/tc2002124_2181.htm

Blackmon, M. H., Kitajima, M., & Polson, P. G. (2003). Repairing usability problems identified by the cognitive walkthrough for the web. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 497–504). doi:10.1145/642611.642698 http://doi.acm.org/10.1145/642611.642698

Blackmon, M., Polson, P., Kitajima, M., & Lewis, C. (2002). Cognitive walkthrough of the web. *Proceedings of the 2002 ACM SIGCHI Conference in Human Factors in Computing Systems*, 463–470. doi:10.1145/503376.503459 http://doi.acm.org/10.1145/503376.503459

Bolchini, D., & Garzotto, F. (2007). Quality of web usability evaluation methods: An empirical study on MiLE+. *Web Information Systems Engineering–WISE 2007 Workshops*, 481–492. doi:10.1007/978-3-540-77010 http://dx.doi.org/10.1007/978-3-540-77010

Bond, D. (2005). Content analysis. *Encyclopedia of Social Measurement*, Elsevier, New York, NY. (pp. 481-485). doi:10.1016/B0-12-369398-5/00030-X http://dx.doi.org/10.1016/B0-12-369398-5/00030-X

Braun, K., Gadney, M., Haughey, M., Roselli, A., Synstelien, D., Walter, T., & Wertheimer, D. (2002). Introduction: Beyond the buzz: The true meaning of usability. In M. Holzschlad & B. Lawson (Eds.), *Usability: The site speaks for itself* (pp. 2–13). Acocks Green, Birmingham, UK: Glasshaus.

Busch, C., De Maret, P. S., Flynn, T., Kellum, R., Le, S., Meyers, B., ..White, R. (2005). Content Analysis. *Writing@CSU. Colorado State University Department of English*. Retrieved May 15, 2012, from http://writing.colostate.edu/guides/research/content/

Castillo, J., & Hartson, H.R. (2007). Remote usability testing methods a la carte. Retrieved from http://eprints.cs.vt.edu/archive/00000946/01/TR-07-05_remotemethods.pdf

Chiew, T., & Salim, S. (2003). Webuse: Website usability evaluation tool. *Malaysian Journal of Computer Science, 16*(1), 47–57. Retrieved from http://mjcs.fsktm.um.edu.my/document.aspx?FileName=199.pdf

Chiou, W-C., Lin, C-C., & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995–2006. *Information & Management, 47*(5–6), 282–290. doi:10.1016/j.im.2010.06.00 http://dx.doi.org/10.1016/j.im.2010.06.002

Christensen, L., & Frøkjaer, E. (2010). Distributed usability evaluation: Enabling large-scale usability evaluation with user-controlled instrumentation. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, 118–127. doi: 10.1145/1868914.1868932 http://doi.acm.org/10.1145/1868914.1868932

Cognitive walkthrough. Wikipedia, (2012). Retrieved on June 9, 2012, from http://en.wikipedia.org/wiki/Cognitive_walkthrough

Cole, O. (2009). Aprobe: A non-intrusive framework for software instrumentation. *OC Systems, Inc,* 1–10. Retrieved from http://www.ocsystems.com/pdf/AprobeTechnologyOverview.pdf

Comer, D., Gries, D., Mulder, M., Tucker, A., Turner, A., Young, P., & Denning, P. (1989). Computing as a discipline. *Communications of the ACM, 32*(1), 9–23. doi: 10.1145/63238.63239 http://doi.acm.org/10.1145/63238.63239

Comparison of Usability Evaluation Methods. Wikipedia. (n.d). Retrieved on April 26, 2012, from http://en.wikipedia.org/wiki/Comparison_of_usability_evaluation_methods

Content. (n.d.) Content analysis: An introduction. University of California-Davis. Retrieved on July 8, 2012 from http://psychology.ucdavis.edu/sommerb/sommerdemo/content/intro.htm.

Cornell University Library. (n.d.). What is an annotated bibliography? Retrieved April 21, 2012, from http://olinuris.library.cornell.edu/ref/research/skill28.htm

Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: SAGE Publications, Inc.

Davis, P., & Shipman, F. (2011). Learning usability assessment models for web sites. *Proceedings of the 16th international conference on intelligent user interfaces* (pp. 195–204). doi:10.1145/1943403.1943433 http://dx.doi.org/10.1145/1943403.1943433

Downing, C., & Liu, C. (2011). Assessing web site usability in retail electronic commerce. *2011 IEEE 35th annual computer software and applications conference (COMPSAC)* (pp. 144-151). doi:10.1109/COMPSAC.2011.26 http://dx.doi.org/10.1109/COMPSAC.2011.26

Dumas. J. (2007). The great leap forward: The birth of the usability profession. *Journal of Usability Studies, 2*(2), 54-60. Retrieved from http://upassoc.org/upa_publications/jus/2007_february/dumas_birth_of_usability_profession.pdf

Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty, *Information & Management, 43* (1), 1-14. doi:10.1016/j.im.2005.01.002 http://dx.doi.org/10.1016/j.im.2005.01.002. Retrieved from http://www.sciencedirect.com/science/article/pii/S0378720605000169

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. *Proceedings of the SIGCHI conference on human factors in computing systems: The CHI is the limit* (pp. 80–87). doi:10.1145/302979.303001 http://doi.acm.org/10.1145/302979.303001

Frøkjaer, E., Hertzum, M., & Hornbaek, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 345–352. doi:10.1145/332040.332455 http://doi.acm.org/10.1145/332040.332455

Hallahan, K. (2001). Improving public relations web sites through usability research. *Public Relations Review, 27*(2), 223–239. doi:10.1016/S0363-8111(01)00082-0

Hartson, H., Castillo, J., Kelso, J., Kamler, J., & Neale, W. (1996). Remote evaluation: The network as an extension of the usability laboratory. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common ground*, 228–235. doi:10.1145/238386.238511 http://doi.acm.org/10.1145/238386.238511

HCI—Human-Computer Interaction. Wikipedia (2012). Retrieved on April 20, 2012, from http://en.wikipedia.org/wiki/Human%E2%80%93computer_interaction

Hewitt, M. (1998). *Carrying out a literature review*. Trent Focus Group. Retrieved from http://ce.uoregon.edu/aim/Capstone07/HewittLitReview.pdf

HHS—U.S. Dept. of Health and Human Services. (2006). *Research-based web design & usability guidelines*. US Government Printing Office. Retrieved from http://www.usability.gov/guidelines/guidelines_book.pdf

Holdford. D. (2008). Content analysis methods for conducting research. *Social and Administrative pharmacy, 4*(2), 173-181. doi:10.1016/j.sapharm.2007.03.003 http://dx.doi.org/10.1016/j.sapharm.2007.03.003

Hollingsed, T., & Novick, D. G. (2007). Usability Inspection methods after 15 years of research. *SIGDOC '07 Proceedings of the 25th annual ACM international conference on design of communication* (pp. 668–668). doi:10.1145/1297144.1297200

Hom, J. (2011). The usability methods toolbox. Retrieved from http://usability.jameshom.com/

Hornbæk, K.,& Law, E.(2007). Meta-analysis of correlations among usability measures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 617-626. doi=10.1145/1240624.1240722 http://doi.acm.org/10.1145/1240624.1240722

Insch, G., Moore, J., & Murphy, L. (1997). Content analysis in leadership research: Examples, procedures, and suggestions for future use. *The Leadership Quarterly*, *8*(1), 1–25. doi: 10.1016/S1048-9843(97)90028-X http://dx.doi.org/10.1016/S1048-9843(97)90028-X

Insfran, E., & Fernandez, A. (2008). A systematic review of usability evaluation in web development. *Web Information Systems Engineering–WISE 2008 Workshops*, 81–91. doi:10.1007/978-3-540-85200-1_10 http://dx.doi.org/10.1007/978-3-540-85200-1_10

Ivory, M., & Hearst, M. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys (CSUR), 33*(4), 470–516. doi:10.1145/503112.503114 http://doi.acm.org/10.1145/503112.503114

Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. (1991). User interface evaluation in the real world: A comparison of four techniques. *Proceedings of the SIGCHI conference on human factors in computing systems: Reaching through technology* (pp. 119–124). doi:http://doi.acm.org/10.1145/108844.108862

Jokela, T., Iivari, N., Matero, J., & Karukka, M. (2003). The standard of user-centered design and the standard definition of usability: Analyzing ISO 13407 against ISO 9241-11. *Proceedings of the Latin American conference on Human-computer interaction,* 53–60.

Keshav, S. (2007). *How to read a paper*. David R. Cheriton School of Computer Science, University of Waterloo, ON, Canada, pp. 1-2. Retrieved from http://ccr.sigcomm.org/online/files/p83-keshavA.pdf.

Kjeldskov, J., Skov, M.B., & Stage, J. (2004). Instant data analysis: Conducting usability evaluations in a day. *Proceedings of the Third Nordic Conference on Human-Computer Interaction* (NordiCHI '04), 233-240. doi: 10.1145/1028014.1028050 http://doi.acm.org/10.1145/1028014.1028050

Krippendorf, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, Ca. Sage Publications. Retrieved from http://books.google.com/books?hl=en&lr=&id=q657o3M3C8cC&oi=fnd&pg=PA3&dq=content+analysis+definition&ots=bKccC3NaxX&sig=g5odXyXxTZ7RbwXiMK4rA1vc8H4#v=onepage&q=content%20analysis%20definition&f=false

Krug, S. (2006*). Don't make me think: Common sense approach to web usability* (2nd ed.). Berkeley, Ca. New Riders Press.

Lallemand, C. (2011). Toward a closer integration of usability in software development: A study of usability inputs in a model-driven engineering process. *Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (EICS '11), 299–302. doi:10.1145/1996461.1996541 http://dx.doi.org/10.1145/1996461.1996541

Lee, Y., & Kozar, K. A. (2012). Understanding of website usability: Specifying and measuring constructs and their relationships. *Decision Support Systems, 52*(2), 450–463. doi:10.1016/j.dss.2011.10.004

Levi, M.D., & Conrad, F.G. (2008). Usability testing of world wide web sites. *U.S. Bureau of Labor Statistics Office of Survey Methods Research.* Retrieved from http://www.bls.gov/ore/htm_papers/st960150.htm

Literature reviews. (2012). *The Writing Center, University of North Carolina at Chapel Hill*. Retrieved May 24, 2012, from http://writingcenter.unc.edu/resources/handouts-demos/specific-writing-assignments/literature-reviews

Lund, A. (2006). Post-modern usability. *Journal of Usability Studies, 2*(2), 54-60. Retrieved from http://upassoc.org/upa_publications/jus/2006_november/lund_post_modern_usability.pdf

Lynch, P., & Horton, S. (2008). *Web style guide: Basic principles for creating web sites* (3rd ed.). New Haven, CT: Yale University Press.

Manzari, L., & Trinidad-Christensen, J. (2006). User-centered design of a web site for library and information science students: Heuristic evaluation and usability testing. *Information*

technology and libraries. Retrieved from http://faculty.mercer.edu/lewis_am/pdfs/UCD%20for%20Library.pdf

Matera, M., Rizzo, F., & Carughi, G. (2008). Web usability: Principles and evaluation methods. *Handbook of Research on Web Information System Quality*, CC Muñoz, MÁ Moraga and M. Piattini, eds., Information Science Reference, IGI Global, Hershey, 234–249. Retrieved from http://webml.elet.polimi.it/webml/upload/ent5/1/WebUsability-MateraEtAl.pdf

Merriam-Webster. (n.d.). Heuristic. Retrieved on April 29, 2012, from http://www.usability.gov/methods/test_refine/heuristic.html

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 373–380. doi:10.1145/142750.142834 http://doi.acm.org/10.1145/142750.142834

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependenc*e *(CHI '94)*, 152-158. doi:10.1145/191666.191729 http://doi.acm.org/10.1145/191666.191729

Nielsen, J. (2000). *Designing* w*eb usability: The practice of simplicity*. Indianapolis, IN: New Riders Publishing.

Nielsen, J. (2003). Usability 101: Introduction to usability. *Jakob Nielsen's Alertbox*, *25*. Retrieved on June 3, 2012, from http://www.hh.se/download/18.5173bcf712de11663378000958/diskussionsuppgift_F5_nielsen.pdf

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *CHI '90 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People*, 249–256. doi:10.1145/97243.9728

NIST—National Institute of Standards and Technology. (2006). Industry usability reporting. Retrieved on April 29, 2012, from http://zing.ncsl.nist.gov/iusr/overview.html

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010). Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination websites for usability. *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, 2381–2390. Retrieved from http://dmtlarchive.skku.edu/share/CHI2010%20Proccedings%20%20/p2381.pdf

Palmquist, R.A. (2001). An overview of usability. *Journal of Education for Library and Information Science, 42*(2), pp. 123-136. Retrieved from http://www.jstor.org/stable/40324025

Protocol analysis. Wikipedia (2012). Retrieved on June 6, 2012, from http://en.wikipedia.org/wiki/Protocol_analysis

Qi, S., Ip, C., Leung, R., & Law, R. (2010). A new framework on website evaluation. *2010 International Conference on E-Business and E-Government (ICEE)*, 78-81. doi:10.1109/ICEE.2010.27 http://dx.doi.org/10.1109/ICEE.2010.27

Redish, J. (Ginny). (2010). Technical communication and usability: Intertwined strands and mutual influences. *IEEE Commentary*. doi:10.1109/TPC.2010.2052861

Redish, J., Bias, R.G., Bailey, R., Molich, R., Dumas, J., & Spool, J. (2002). Usability in practice: Formative usability evaluations - evolution and revolution. In *CHI '02 extended*

abstracts on Human factors in computing systems (CHI EA '02). ACM, New York, NY, USA, doi:10.1145/506443.506647 http://doi.acm.org/10.1145/506443.506647

Rosson, M., Carroll, J., & Rodi, C. (2004). Case studies for teaching usability engineering. *ACM SIGCSE Bulletin*, 36, 36–40. doi:10.1145/1028174.971315

Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing*. Wiley Publishing, Inc. Indianapolis, IN.

Seffah, A., & Metzker, E. (2004). The obstacles and myths of usability and software engineering. *Communications of the ACM, 47*(12), 71–76. doi:10.1145/1035134.1035136 http://dx.doi.org/10.1145/1035134.1035136

Skidmore College. (n.d.) Retrieved from http://lib.skidmore.edu/library/index.php/li371-annotated-bib

Spool, J., (2009). The $300 million button. *User Interface Engineering*. Retrieved from http://www.uie.com/articles/three_hund_million_button/

Tedeschi, B. (1999, August 30). E-commerce report: Online merchants find that a well-designed web site can have a big impact on bottom line. *New York Times*. Retrieved May 30, 2012, from http://www.nytimes.com/1999/08/30/business/e-commerce-report-line-merchants-find-that-well-designed-web-site-can-have-big.html?pagewanted=print&src=pm

Teo, H.H., Oh, L.B., Liu, C., & Wei, K.K. (2003). An empirical study of the effects of interactivity on web user attitude. *International Journal of Human-Computer Studies, 58*(3), 281–305. doi:10.1016/S1071-5819(03)00008-9 http://dx.doi.org/10.1016/S1071-5819(03)00008-9

Tullis, T., & Albert, W. (2008). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann Publishers. Retrieved from https://www10.iiba.org/source/IIBA_24x7/index.cfm?Section=Learning

University of California, Berkeley Library, n.d., Retrieved April 29, 2012 from http://www.lib.berkeley.edu/instruct/guides/evaluation.html

UPA—Usability Professionals' Association. (n.d.) What is user centered design? Retrieved April 29, 2012, from http://www.usabilityprofessionals.org/usability_resources /about_usability/what_is_ucd.html

Usability.gov (n.d) Usability Basics. Retrieved on April 29, 2012 from http://www.usability.gov/basics/index.html

Usability.gov (n.d.). Usability evaluations. Retrieved on April 29, 2012, from http://www.usability.gov/methods/test_refine/learneval.html

Usability.gov (n.d.) User-Centered Design. Retrieved on April 29, 2012, from http://www.usability.gov/basics/ucd/index.html

Usability.gov (n.d.). Usability Testing. Retrieved on April 29, 2012, from http://www.usability.gov/methods/test_refine/learnusa/index.html

Usability BOK (Body of Knowledge). (n.d.). Retrieved on June 11, 2012, from http://www.usabilitybok.org/glossary.

UseIt.com. (n.d). Heuristic Evaluation. Retrieved on April 29, 2012, from http://www.useit.com/papers/heuristic/

User Experience. Wikipedia (2012). Retrieved April 16, 2012, from http://en.wikipedia.org/wiki/User_experience

UXnet. *User Experience Network (2010)*. Retrieved April 10, 2012, from http://www.uxnet.org/

Whitehead, C. (2006). Evaluating web page and web site usability. *Proceedings of the 44th annual Southeast regional conference*, (pp. 788–789). doi:10.1145/1185448.1185637 http://dx.doi.org 10.1145/1185448.1185637

Wixon, D. (2011). The unfulfilled promise of usability engineering. *Journal of Usability Studies*, *6*(4), 198–203. Retrieved from http://www.upassoc.org/upa_publications/jus/2011august/JUS_Wixon_August_2011.pdf