

THE USABILITY OF DIGITAL LIBRARY LEARNING RESOURCES: DEVELOPER
PERCEPTIONS OF THE RESULTS OF AN EMPIRICAL AUTOMATED USABILITY
EVALUATION APPROACH

by

SAUN SHEWANOWN

(Under the Direction of Thomas C. Reeves)

ABSTRACT

This two-part study focused first on gaining insight into the current state of Web-based learning site usability, and second into better understanding of the perceptions of Web-based learning site builders regarding the usefulness of usability analysis results generated by an automated tool. Part one of this study involved analysis of usability attributes of Web-based learning sites using an empirically-based automated Web usability evaluation tool. Part two of the study involved Web-based surveys focused on the perceptions of Web-based learning site builders about Web usability. A literature review focused on the topics of usability, Web usability, Web-based learning, Web-based learning usability evaluation, and automated Web site usability evaluation tools supports the rationale for this study.

To begin the study, an automated usability evaluation, using a software program called WebTango developed at the University of California at Berkeley (UCB), was performed on Web sites that are part of the DLESE Reviewed Collection (DRC), a subset of the educational resources in the Digital Library for Earth System Education (DLESE).

Subsequently, a Web-based cross sectional survey was distributed to the builders of a subset of the DLESE resources that had been subjected to the usability evaluation. The first part of the survey explored the builders' perceptions of Web usability in general, and the second part explored their perceptions of the analysis results yielded by WebTango with respect to the particular Web-based learning sites developed by the builders. Both Likert Scale (agree-disagree) and open-ended response formats was used for description and exploration.

The results of the first part of the study indicated that the usability quality of Web-based learning sites was rated by WebTango as average or below average. In the survey study that constituted the second part of the study, respondents (builders of DLESE resources) were generally unfamiliar with usability and had little expertise or experience in applying usability evaluation methods. They expressed a desire to learn more about usability and enhance the usability of their resources, but they were generally uncertain regarding the utility and value of an automated evaluation tool. The respondents identified some benefits and issues of an automated evaluation tool that may be helpful for future research and development of automated usability evaluation tools.

INDEX WORDS: human-computer interaction, HCI, Web-based learning, educational Web sites, learning Web sites, e-learning, eLearning, usability, WebTango, automated usability evaluation

THE USABILITY OF DIGITAL LIBRARY LEARNING RESOURCES: DEVELOPER
PERCEPTIONS OF THE RESULTS OF AN EMPIRICAL AUTOMATED USABILITY
EVALUATION APPROACH

by

SAUN SHEWANOWN

A.B., The University of California at Berkeley, 1992

M.S., Southern Polytechnic State University, 1998

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2006

© 2006

Saun Shewanown

All Rights Reserved

THE USABILITY OF DIGITAL LIBRARY LEARNING RESOURCES: DEVELOPER
PERCEPTIONS OF THE RESULTS OF AN EMPIRICAL AUTOMATED USABILITY
EVALUATION APPROACH

by

SAUN SHEWANOWN

Major Professor: Thomas C. Reeves

Committee: Lloyd P. Rieber
Ikseon Choi
Norman Thomson

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2006

Dedication

To Marsha.

ACKNOWLEDGEMENTS

My doctoral study journey started when I was teaching computer science at SPSU. I am fortunate to have crossed paths with many wonderful people. Dr. Mike Murphy encouraged and advised me on starting my pursuit of a doctorate. He allowed my teaching schedule to accommodate my course schedule at UGA. Dr. Evan Glazer mentored me and allowed me to stay at his house when I visited Athens. My personal experience with Dr. Mike Hannafin and Dr. David Gries was inspirational. I am grateful that Dr. Janet Kolodner and Dr. Mark Guzdail allowed me to observe their research projects. I am especially thankful for Dr. Kolodner, Dr. Gries, and Mike McCracken, who spent a great deal of their valuable time to advise me individually. Looking back, if I had taken some of the practical advice that Mike McCracken had given, my doctoral study would have gone much smoother.

I would like to thank my committee members—Dr. Thomas C. Reeves, Dr. Lloyd Reiber, Dr. Norm Thompson, and Dr. Ikseon Choi—who were willing to lift me up from the dust and guide me toward the successful completion of the degree. I also would like to thank Shannon Whitehead and Lisa Grimenstein for their hard work, patience, and responsiveness in editing the drafts. I appreciate Christy Glazer and Cathy Dickson for taking care of all the administrative necessities. Thanks to the students enrolled in the Fall 2005 Instructional Product Evaluation (EDIT 8350) course at the University of Georgia for performing a test-run of the survey Web site. I especially appreciate Chun-Min Wang, a fellow UGA doctoral student, who assisted me in analyzing the sites using WebTango. I am grateful for the generous support of the survey participants.

I am truly thankful for the continuing friendship, moral support, and genuine concern of Bill Lahey, Dr. Sabastian Loh, and Dr. Yuxin Ma. A thank you goes to Alvin and Sarah Levy, Mr. and Mrs. Pak Pang Chow, Jack and Eva Chang, Linda Ware, Tom Demetriou, Seungwon Yang, Yannick Bashizi, Amit Thuthija, Robert Persaud, Seungyeon Han, Kuan-Chung Chen, Ying Liu, David Brill, Rosemary Lynn, Geri Corcoran, Michele Estes, Christa Harrelson, Greg Clinton, Marty Thomas, Xornam Apedoe, Jo McClendon, Liyan Song, and Dr. Janette R. Hill for their constant well-wishes. You have been part of my journey.

Dr. Reeves—my advisor, teacher, role model, and a wonderful human being—has embraced me when I was struggling and nurtured my intellectual growth. He genuinely cares about my well-being. I am still amazed at his ability to “make time”. Words can’t describe my gratitude toward Dr. Reeves.

My wife Marsha has been consistently supportive of my pursuit of education. She not only stood by me throughout my doctoral study, but also my undergraduate study at U.C. Berkeley—both challenging, although in different ways. Marsha has celebrated the good times and endured the hard times with me. I can’t imagine getting to this point in my education without Marsha.

I am very grateful to be given the opportunities to bring me this far in my formal education. I believe that each milestone of this journey is the mark of a beginning, not an “accomplishment”. I sincerely hope that this particular milestone marks a new beginning in my continued pursuit of this vision: making a positive contribution to my surroundings.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	V
TABLE OF CONTENTS.....	VII
LIST OF TABLES.....	X
LIST OF FIGURES	XIII
Chapter 1: Introduction.....	1
Web-based Learning.....	1
Educational and Usability Considerations of Web-based Learning	3
Usability Issues in Web-based Learning.....	4
The Goals and Research Questions of this Study	6
Assumptions and Limitations	7
Chapter 2: Literature Review.....	9
Usability and Web Site Usability.....	10
Automated Web Site Usability Evaluation Methods and Tools	27
WebTango.....	32
WebTango Related Research Studies	39
Web-based Learning	51
Web-based Learning Usability Related Studies	56
Conclusion	63
Chapter 3: Methodology	65

Phase I: Web-based Learning Site Usability Evaluation	66
Phase II: Survey the Web-based Learning Web Site Developers.....	71
Summary	76
Chapter 4: The Pilot Survey.....	78
The Heuristic Evaluation	78
Results of Research Questions #2.1 and #2.2	79
Results of Research Question #2.3	82
Findings of Research Questions #3.....	86
Summary	90
Chapter 5: Results	92
Results of Research Question #1	93
Sites Included in the Survey Study	96
Results of Research Questions #2.1, and #2.2	101
Results of Research Questions #2.3.....	106
Findings of Research Question #3	111
Limitations	118
Summary	119
Chapter 6: Conclusions and Discussion.....	120
Study Overview	120
Conclusions.....	122
Implications.....	128
Future Research Directions.....	132

Summary	136
REFERENCES	138
APPENDICES	154
Appendix A: Automated Guideline Review Usability Evaluation Tools.....	154
Appendix B: The WebTango Web Interfaces Usability Measurements.....	155
Appendix C: WebTango Tools	162
Appendix D: Sample Excluded Sites.....	165
Appendix E: Survey Instruments.....	167
The First Instrument: Addressing Research Questions #2.1, #2.2, and #2.3	167
The Second Instrument: Addressing Research Question #3	171
Appendix F: The Pilot Study Results Histograms	180
The First Survey Instrument: Addressing Research Questions #2.1 and #2.2.....	180
The First Survey Instrument: Addressing Research Questions #2.3	184
The Second Survey Instrument: Addressing Research Question #3	188
Appendix G: The Results Chapter Histograms.....	191
The First Survey Instrument: Addressing Research Questions #2.1 and #2.2.....	191
The First Survey Instrument: Addressing Research Question #2.3	195
The Second Survey Instrument: Addressing Research Question #3	199
Appendix H: Diagram Symbols.....	202

LIST OF TABLES

Table 2-1 Usability Evaluation Approaches	28
Table 2-2 Usability Evaluation Activities Shared by Evaluation Approaches	29
Table 2-3 Types of Evaluation Tools Supporting Evaluation Activities	29
Table 3-1 Usability Evaluation Items Supported by WebTango and Included in this Study	67
Table 4-1 The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #1 through #6 of the First Survey Instrument in the Pilot Study	81
Table 4-2 The Descriptive Statistics of Survey Questions #1 through #6 of the First Survey Instrument in the Pilot Study	82
Table 4-3 The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #7 through #12 of the First Survey Instrument in the Pilot Study in the pilot study	83
Table 4-4 The Descriptive Statistics of Survey Questions #7 through #12 of the First Survey Instrument in the Pilot Study	84
Table 4-5 The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #1 through #4 of the Second Survey Instrument in the pilot study	87
Table 4-6 The Descriptive Statistics of Survey Questions #1 through #4 of the Second Survey Instrument in the Pilot Study	88
Table 5-1 Evaluation Rating Equivalence	94

Table 5-2	The Number of Ratings (N), Mean, and Standard Deviation of the Eight WebTango Measures of the 130 Sites Included in Phase I of this Study	94
Table 5-3	The Number of Ratings (N), Mean, and Standard Deviation of the Eight WebTango Measures of 35 Sites Included in Phase II	97
Table 5-4	The Descriptive Statistics of ratings. Group 1 being all 130 sites. Group 2 being the 35 Surveyed Sites	98
Table 5-5	The t-test Results for Comparing the Two Groups	100
Table 5-6	The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #1 through #6 of the First Survey Instrument	103
Table 5-7	The Descriptive Statistics of Survey Questions #1 through #6 of the First Survey Instrument	104
Table 5-8	The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #7 through #12 of the First Survey Instrument	107
Table 5-9	The Descriptive Statistics of Survey Questions #7 through #12 of the First Survey Instrument	108
Table 5-10	The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #1 through #4 of the Second Survey Instrument	112
Table 5-11	The Descriptive Statistics of Survey Questions #1 through #4 of the Second Survey Instrument	113

Table 6-1	The Four Combinations of Usability Knowledge Level and Web site Usability	
	Quality Level	131
Table A-1	Automated Guideline Review Usability Evaluation Tools (Ivory, 2003). ..	154
Table B-1	The 157 Quantitative Measures of WebTango for Evaluating Web Interfaces	
	(Ivory, 2001).	155
Table B-2	WebTango Measures: Graphic Elements (Ivory, 2001).	156
Table B-3	WebTango Measures: Graphic Formatting (Ivory, 2001).	156
Table B-4	WebTango Measures: Link Elements (Ivory, 2001).	156
Table B-5	WebTango Measures: Link Formatting (Ivory, 2001).....	156
Table B-6	WebTango Measures: Page Formatting (Ivory, 2001).	157
Table B-7	WebTango Measures: Page Performance (Ivory, 2001).....	158
Table B-8	WebTango Measures: Site Architecture (Ivory, 2001).....	159
Table B-9	WebTango Measures: Text Elements (Ivory, 2001).....	160
Table B-10	WebTango Measures: Text Formatting (Ivory, 2001).....	161

LIST OF FIGURES

Figure 2.1. Sample keywords used for locating literature related to this dissertation.	10
Figure 2.2. A usability test: User(s) perform predetermined tasks while being observed and recorded.	14
Figure 2.3. The six phases of a formal usability testing process.	15
Figure 2.4. The usability testing process and the deliverables of each phase.	17
Figure 2.5. The dimensions of usability evaluation approaches, activities, and tools (Ivory, 2003). Automated guideline review , highlighted, was the focus of this dissertation which will be discussed in more detail in section Automated Guideline Review Tools.	28
Figure 2.6. WebTango Analysis tool and TangoViewer tool screen samples.	33
Figure 2.7. WebTango Evaluation process overview.	33
Figure 2.8. A sample WebTango page-level prediction model (good, average, and poor) (Ivory, 2001, p. 156).	35
Figure 2.9. WebTango page-level measurement model development process.	37
Figure 2.10. WebTango site-level measurement model development process.	38
Figure 2.11. An overview of Linking Web Interface Profiles to Usability study.	40
Figure 2.12. An overview of the Evaluating the Web Interface Profiles study.	43
Figure 2.13. An overview of A Study of Automated Web Site Evaluation Tools.	44
Figure 2.14. The overview of Clapsaddle’s methodology (Clapsaddle, 2004).	49
Figure 2.15. A decision tree model and rule set for “good” e-commerce Web sites (Clapsaddle, 2004, p. 88).	50

Figure 2.16. The top measures (Clapsaddle, 2004, p. 83).	50
Figure 2.17. Wang's Level 1 Web-based learning.	53
Figure 2.18. Wang's Level 2 Web-based learning.	54
Figure 2.19. Wang's Level 3 Web-based learning.	54
Figure 3.1. An overview of methodology.	65
Figure 3.2. A screen sample of DLESE (DLESE, 2005b).	68
Figure 3.3. The methodology summary of phase I.	70
Figure 3.4. An overview of the survey study.	72
Figure 4.1. Survey questions #1 through #6 that addressed research questions #2.1 and #2.2.	80
Figure 4.2. The mean scores of survey questions #1 through #6 of the first survey instrument in the pilot study.	81
Figure 4.3. Questions #7 through #12 of the first survey instrument which addressed research question #2.3.	83
Figure 4.4. The average scores of survey questions #7 through #12 of the first survey instrument in the pilot study.	84
Figure 4.5. Questions #1 through #4 of the second survey instrument which addressed research question #3.	86
Figure 4.6. The average scores of questions #1 through #4 of the second survey instrument in the pilot study.	88
Figure 5.1. The frequency distribution of the eight WebTango measures of the 130 sites included in Phase I of this Study.	95
Figure 5.2. The frequency distribution of the eight WebTango measures of 35 sites.	97

Figure 5.3 The trends of the means of the two groups: Group 1 being all 130 sites. Group 2 being the 35 surveyed sites.	99
Figure 5.4. Survey questions #1 through #6 that address research questions 2.1 and 2.2.	102
Figure 5.5. The average expertise and use of usability methods.	103
Figure 5.6 The mean scores (of questions #1 through #6) comparisons between the actual survey and the pilot survey.	105
Figure 5.7 The average expertise and use of usability methods comparisons between the actual survey and the pilot survey.....	105
Figure 5.8. Questions #7 through #12 of the first survey instrument which addressed research question 2.3.....	106
Figure 5.9. The average scores of survey questions #7 - #12 of the first survey instrument.	107
Figure 5.10 The mean scores (of questions #7 through #12) comparisons between the actual survey and the pilot survey.....	108
Figure 5.11. Questions #1 through #4 of the second survey instrument which address research question #3.....	111
Figure 5.12. The average scores of questions #1 through #4 of the second survey instrument.	112
Figure 5.13 The mean scores (of questions #1 through #4) comparisons between the actually survey and the pilot survey.....	113
Figure C.1. WebTangle Analysis tool screen sample (Ivory-Ndiaye, 2004b).....	162
Figure C.2. TangoViewer tool screen sample (Ivory-Ndiaye, 2004a).....	162

Figure C.3. WebTango initial setup process.....	163
Figure C.4. WebTango analysis process.....	164
Figure D.1. A site that is a description of an actual learning material.....	165
Figure D.2. Learning material that is part of a large portal.	166
Figure E.1. The invitation for the first survey instrument.	167
Figure E.2. The informed consent form for the first survey instrument.	168
Figure E.3. Survey questions #1 through #6 addressing research questions #2.1 and #2.2.	169
Figure E.4. Survey questions #7 through #12 addressing research question #2.3.....	169
Figure E.5. The usability evaluation method descriptions accompanying survey questions #7 through #12.	170
Figure E.6. A sample WebTango analysis results.	171
Figure E.7. Sample text results summary.	172
Figure E.8. Sample graphic results summary.	173
Figure E.9. Sample link results summary.	174
Figure E.10. Sample page results summary.....	175
Figure E.11. A description of how to read the evaluation results summary.....	176
Figure E.12. The invitation for the second survey instrument.....	177
Figure E.13. The informed consent form for the second survey instrument.	178
Figure E.14. The second survey instrument addressing research question #3.....	179
Figure F.1. Questions #1 through #6 of the first survey instrument which addressed research questions #2.1 and #2.2.	180

Figure F.2. The central tendency, variability, size, and distribution of question #1 of the first survey instrument.	181
Figure F.3. The central tendency, variability, size, and distribution of question #2 of the first survey instrument.	181
Figure F.4. The central tendency, variability, size, and distribution of question #3 of the first survey instrument.	182
Figure F.5. The central tendency, variability, size, and distribution of question #4 of the first survey instrument.	182
Figure F.6. The central tendency, variability, size, and distribution of question #5 of the first survey instrument.	183
Figure F.7. The central tendency, variability, size, and distribution of question #6 of the first survey instrument.	183
Figure F.8. Questions #7 through #12 of the first survey instrument which addressed research question #2.3.....	184
Figure F.9. The central tendency, variability, size, and distribution of question #7 of the first survey instrument.	185
Figure F.10. The central tendency, variability, size, and distribution of question #8 of the first survey instrument.	185
Figure F.11. The central tendency, variability, size, and distribution of question #9 of the first survey instrument.	186
Figure F.12. The central tendency, variability, size, and distribution of question #10 of the first survey instrument.	186

Figure F.13. The central tendency, variability, size, and distribution of question #11 of the first survey instrument.	187
Figure F.14. The central tendency, variability, size, and distribution of question #12 of the first survey instrument.	187
Figure F.15. Questions #1 through #4 of the second survey instrument.	188
Figure F.16. The central tendency, variability, size, and distribution of question #1 of the second survey instrument.	188
Figure F.17. The central tendency, variability, size, and distribution of question #2 of the second survey instrument.	189
Figure F.18. The central tendency, variability, size, and distribution of question #3 of the second survey instrument.	189
Figure F.19. The central tendency, variability, size, and distribution of question #4 of the second survey instrument.	190
Figure G.1. Questions #1 through #6 of the first survey instrument which addressed research questions #2.1 and #2.2.	191
Figure G.2. The central tendency, variability, size, and distribution of question #1 of the first survey instrument.	192
Figure G.3. The central tendency, variability, size, and distribution of question #2 of the first survey instrument.	192
Figure G.4. The central tendency, variability, size, and distribution of question #3 of the first survey instrument.	193
Figure G.5. The central tendency, variability, size, and distribution of question #4 of the first survey instrument.	193

Figure G.6. The central tendency, variability, size, and distribution of question #5 of the first survey instrument.	194
Figure G.7. The central tendency, variability, size, and distribution of question #6 of the first survey instrument.	194
Figure G.8. Questions #7 through #12 of the first survey instrument which addressed research questions #2.3.	195
Figure G.9. The central tendency, variability, size, and distribution of question #7 of the first survey instrument.	195
Figure G.10. The central tendency, variability, size, and distribution of question #8 of the first survey instrument.	196
Figure G.11. The central tendency, variability, size, and distribution of question #9 of the first survey instrument.	196
Figure G.12. The central tendency, variability, size, and distribution of question #10 of the first survey instrument.	197
Figure G.13. The central tendency, variability, size, and distribution of question #11 of the first survey instrument.	197
Figure G.14. The central tendency, variability, size, and distribution of question #12 of the first survey instrument.	198
Figure G.15. Questions #1 through #4 of the second survey instrument.....	199
Figure G.16. The central tendency, variability, size, and distribution of question #1 of the second survey instrument.....	199
Figure G.17. The central tendency, variability, size, and distribution of question #2 of the second survey instrument.....	200

Figure G.18. The central tendency, variability, size, and distribution of question #3 of the second survey instrument.....	200
Figure G.19. The central tendency, variability, size, and distribution of question #4 of the second survey instrument.....	201

Chapter 1: Introduction

The Web has increasingly become an integral part of teaching and learning in higher education since its beginnings a dozen years ago (Khan, 1997). Pedagogy is necessarily the paramount consideration when building a Web-based learning environment, but usability is another critically important issue. Although many research studies have been conducted about the pedagogical aspects of Web-based learning in higher education (Bruning, Horn, & PytlikZillig, 2003; Jonassen, 2004), usability is a less visible topic in the educational research literature. The use of automated usability evaluation tools in designing and evaluating Web-based learning environments is an especially under-explored topic. Accordingly, this dissertation focused on the implications of using automated usability evaluation tools to enhance the design and utility of Web-based learning environments. This chapter is intended to describe the nature of Web-based learning in higher education, specify usability issues related to Web-based learning, and clarify the goals of this dissertation.

Web-based Learning

Web-based learning (WBL) has become so widely utilized in higher education that many students have come to expect it to be a component of their courses almost as much as textbooks and other traditional resources (Bruning et al., 2003). Storey, Phillips, Maczewski, and Wang (2002) found that, even in traditional brick and mortar educational institutions, students expect Web-based course supplements as part of their educational experience. Distance-education programs and courses employing the Web as a delivery system are offered by even the most reputable institutions (Carnevale, 2004a; CMU, 2005; eCornell, 2005). Distance and flexible education course offerings and enrollments

have expanded in postsecondary education institutions in recent years, largely due to the ever more sophisticated affordances of the Web (Waits & Lewis, 2003). Whereas many instructors have rudimentary Web sites containing a syllabus and perhaps a few links, some professors use interactive resources available from external online sources (e.g., <http://www.merlot.org/>) to supplement their own courses and a few even use whole online courses developed by others (Carnevale, 2004b). Free or low cost digital educational resources are increasingly accessible through digital libraries (Arms, 2000).

Reasons abound for postsecondary education institutions to expand their distance and flexible education programs via the Web, including providing higher education opportunities to students who would otherwise not have them and enhancing the quality of teaching and learning. These trends are even extending into the K-12 sector where schools benefit by:

offering courses not otherwise available at the school [e.g., AP or college courses],... addressing growing populations and limited space, reducing scheduling conflicts for students, permitting students who failed a course to take it again, meeting the needs of specific groups of students, and generating more district revenues (Setzer & Lewis, 2005, p. 14).

In higher education, Web-based learning takes many forms, ranging from simply enhancing traditional face-to-face courses with online resources to whole degree programs that require students and instructors to meet entirely in the virtual spaces of the Web. To help meet the needs of the various educators requiring Web-based resources, the Massachusetts Institute of Technology (MIT) is just one of several institutions that has developed software for creating and running online learning courses and has distributed this software and related resources for free (Carnevale, 2004a). Despite the availability of free or low cost open source tools to support Web-based learning, most

colleges and universities have invested in some form of commercial course management system software such as WebCT or BlackBoard for their faculty. Meanwhile, individual faculty members, some academic departments, and a few entire institutions are using open source software programs such as Moodle and ATutor to support web-based learning.

Educational and Usability Considerations of Web-based Learning

As noted above, building an effective Web-based learning environment should involve *pedagogical* considerations first and foremost, but usability should not be overlooked. *Pedagogical* considerations, such as learning principles, instructional design, human cognition, and human social interaction principles are essential to achieve learning outcomes (Barab, MaKinster, Moore, & Cunningham, 2001; Clark & Mayer, 2003; Kirschner, Strijbos, Kreijns, & Beers, 2004; Mayer, 2001; Schank & Cleary, 1995). Usability considerations are also necessary for building a Web-based learning environment to achieve learning, although not sufficient on their own (Kirschner et al., 2004).

The importance of usability in WBL was emphasized by Wang (2003) who wrote:

Interface design flaws can cause severe problems when students use computer tools in classroom. In the preliminary phase of interface design, designers should adhere to the standards for usability and accessibility... conducting usability testing with representatives of the end users is essential to eliminating potential operational problems in the future. The cognitive load demanded by an interface should be minimized so that students can focus their cognitive processing on the learning tasks (p. 136).

Kirschner et al. (2004) also suggested

Neglecting usability criteria risks creating [Web-based] CSCL environments that contain all the needed educational and social functionalities (in Nielsen's 1994 terminology utility), but that cannot be

handled by their users (i.e., the learners) because they are difficult to learn, access, and/or control (p. 50).

Storey et al. (2002) concluded that Web-based learning tools with poor usability may not be used at all, and that participants would be in favor of using a Web-based learning tool only if the tool was functional and easy to use. Miller (2005) has succinctly summed-up the impact of usability on e-learning:

While the usability and educational effectiveness of an e-learning application are not one and the same, the two arguably have very much in common. Even though many organizations have made great strides in their ability to develop and deliver e-learning programs to their employees, customers, and suppliers, the usability of these e-learning applications is often lacking or entirely overlooked. Given the large investments organizations are making in online training, and the unique needs of learners, it would be prudent to address the usability of e-learning applications. Doing so will help ensure that users can actually access the necessary material, have optimal levels of satisfaction with the learning experience, and enable the organization to maximize its e-learning investment. (Bottom Line section, ¶ 1)

Usability Issues in Web-based Learning

Similar to Web sites in general, most educational Web sites have poor usability quality (Kirschner et al., 2004). Storey et al. (2002) evaluated two commercially available, widely used Web-based learning systems for educational institutions, finding they violated most of the established usability principles (Norman & Draper, 1986). This violation of usability principles “negatively impacted students and their attitudes towards these tools” (Storey et al., 2002, p. 92).

Usability testing should be part of the formative evaluation processes applied during the development of a Web-based learning environment (Hughes & Burke, 2001; Reeves & Hedberg, 2003). Ideally, a usability test of Web-based training should be conducted in a usability lab by a multi-disciplinary team that includes all stakeholders, such as learners, instructional designers, Web developers and programmers, training and

development managers, and risk managers (Hughes & Burke, 2001). A usability testing process should include various activities: a planning meeting, a scenarios generation session, a walk-through, rehearsal, assessment session, and an action meeting (Hughes & Burke, 2001). When building a Web-based learning site, usability evaluations should be conducted early and often, including both heuristic evaluation and formal usability testing (Reeves & Hedberg, 2003).

Unfortunately, conducting a formal usability test can be time-consuming and resource intensive (Reeves & Carter, 2001). People resist formative evaluation activities such as usability testing for various reasons: lack of time, lack of money, human nature, unrealistic expectations, measurement difficulties, and lack of knowledge in evaluation (Flagg, 1990). When conducting a formal usability test is not feasible, other usability evaluation methods, such as heuristic evaluation (Nielsen, 1993), may be used to supplement or substitute for a usability test. Although a heuristic evaluation is less comprehensive than a usability test, using a heuristic evaluation as a supplement may reduce the scope, time, and resources required for a formal usability test by identifying potential issues prior to conducting the usability test (Ivory & Hearst, 2001; Kjeldskov, Skov, & Stage, 2004; Miller, 2005; Reeves & Hedberg, 2003).

Certainly, conducting any usability evaluation is better than conducting no test at all. An automated usability evaluation is another method that can be used to supplement a usability test when human, time, and financial resources are lacking. In the commercial world, supplementing traditional usability methods with automated usability evaluation tools to improve Web site usability has led to discovering more usability issues and saving time and resources (Brajnik, 2004; Clapsaddle, 2004; Elizabeth & Tharam, 1997;

Forsythe, 2003; Ivory, 2001, 2003; Laskowski, Landay, & Lister, 2002; Obendorf, Weinreich, & Hass, 2004; Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004). However, automated usability tools have not been applied to online educational resources per se. Accordingly, this dissertation has explored the use of automated usability evaluation tools to evaluate the usability of Web-based learning sites.

In summary, the Web has increasingly become an essential part of teaching and learning environments, especially in higher education. Pedagogy and usability are two critically important factors in effective Web-based learning. Although high usability quality is not sufficient to address learning effectiveness, it should not be ignored. The usability quality of Web-based learning is generally low. The need to examine the utility of automated usability evaluation approaches in the context of Web-based learning resources is clear, and this study was designed to do so.

The Goals and Research Questions of this Study

This dissertation has been carried out to achieve two major goals: to gain insight into the current state of Web-based learning site usability based on an automated evaluation tool, and to understand the perceptions of Web-based learning site builders on the usefulness of usability analysis results generated by an automated tool. First, this study involved the analysis of the usability quality of Web-based learning resources using WebTango (Ivory-Ndiaye, 2004a, 2004b), an empirically-based automated usability evaluation tool, and second, this study involved a two-part survey of the creators of Web-based learning resources to reveal their understanding of usability issues in general and their perceptions of the analysis results of WebTango in particular. The following research questions guided this study:

1. What is the usability quality of Web-based learning sites according to an automated usability evaluation tool?
2. According to the perception of Web-based learning site builders,
 - 2.1. What is the degree of expertise among Web-based learning site builders regarding usability evaluation methods?
 - 2.2. How frequently do the Web-based learning site builders use usability evaluation methods?
 - 2.3. What is the *perception* among Web-based learning site builders regarding usability evaluation?
3. According to the perception of Web-based learning site builders, how useful are the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality?

Assumptions and Limitations

The assumptions for this study are as follows: First, Web-based learning has become an increasingly important component of higher education (Bruning, Horn, & PytlikZillig, 2003; Setzer & Lewis, 2005; Storey, Phillips, Maczewski, & Wang, 2002). Second, Web usability is critical (although not sufficient alone) to the effectiveness of Web-based learning (Kirschner, Strijbos, Kreijns, & Beers, 2004; Miller, 2005; Storey, Phillips, Maczewski, & Wang, 2002; F.-K. Wang, Moore, Wedman, & Shyu, 2003). Third, Web site builders typically lack the required resources to conduct formal Web usability testing on an extensive or even sufficient basis (Flagg, 1990; Kjeldskov, Skov, & Stage, 2004; Miller, 2005; Reeves & Hedberg, 2003). Finally, the use of automated Web usability evaluation tools is a viable supplement to formal Web usability testing in

the commercial context (Brajnik, 2004; Clapsaddle, 2004; Elizabeth & Tharam, 1997; Forsythe, 2003; Ivory, 2001, 2003; Laskowski, Landay, & Lister, 2002; Obendorf, Weinreich, & Hass, 2004; Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004).

The limitations of this study are as follows: First, this study focused on usability, not accessibility. Second, this study covered a subset of Web-based learning sites from only one digital library. Third, only one Web usability evaluation tool was used to analyze Web-based learning sites. Finally, the analysis results of Web-based learning sites inherited the limitations of the particular automated Web evaluation tool, WebTango, that was used in the study. Other limitations relevant to each chapter are discussed at the end of that particular chapter.

Chapter 2: Literature Review

This literature review chapter encompasses the following topics: usability in general and Web usability in particular; Web-based learning in general and Web-based learning usability evaluation specifically, and automated Web site usability evaluation tools in general, and WebTango specifically.

The questions this review sought to answer were: What is usability? What is Web usability? What is Web-based learning (WBL)? Why is usability an important issue in WBL? What research has been done to examine the usability of WBL? What is the state-of-the-art of automated Web site usability tools? How has a specific automated tool, WebTango, been used in research examining the Web usability?

The research literature for this review was identified using online resources to locate the research literature most relevant to this dissertation: GALIEO database (GALILEO, 2005), Electronic Journal Locator (University of Georgia Libraries, 2005), ACM Digital Library (ACM, 2005), NetLibrary (NetLibrary, 2005), Google Scholar (Google, 2005b), and Google (Google, 2005a). Figure 2.1 illustrates sample keywords used for this literature review.

This literature review began several years ago in that as a doctoral student in instructional technology, usability and related topics such as graphical user interface (GUI) and human computer interaction (HCI) have long been of interest to me. The literature review reported in this study includes the most recent reports of relevant studies available as of December 2005 when this study was completed.

<ul style="list-style-type: none"> • automated usability evaluation • cognitive walkthrough • distance education • distance learning • ease-of-use • educational Web site • eLearning • e-learning • field observation • focus groups • formal usability test • guideline review • HCI • heuristic evaluation • human-computer interaction • online education • online learning • phenomenological interview • qualitative research • quantitative research • questionnaire • semi-structured interview • structured interview • Survey • survey research 	<ul style="list-style-type: none"> • think-aloud • usability • usability evaluation • usability inquiry • usability inspection • usability issue • usability problem • usability test • usability testing • user friendly • walkthrough • walk-through • WBL • Web evaluation • Web usability • Web usability evaluation • Web usability issue • Web usability problem • Web-based education • Web-based learning • Web-based learning environment • Web-based learning system • Web-based training • WebTango
--	--

Figure 2.1. Sample keywords used for locating literature related to this dissertation.

Usability and Web Site Usability

Usability

Usability is defined 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” (ISO, 1998). *Usability* is closely related to *human-computer interaction* (HCI) and *accessibility*. *Human-computer interaction* “is concerned with understanding how people make use of devices and systems that incorporate or embed

computation, and how such devices and systems can be more useful and more usable” (Carroll, 2003, p. 1). *Accessibility* addresses “issues associated with designing accessible software for people with the widest range of visual, hearing, motor and cognitive abilities, including those who are elderly and temporarily disabled” (ISO, 2003). Although there is a general perception that *usability*, *accessibility*, and *HCI* share some commonalities, each has its own emphasis and scope, and the boundaries of *usability*, *HCI* and *accessibility* are not as sharply defined as might be desired. The following explanations may help to clarify how the terms relate to one another:

- *Usability* is part of *HCI*: "to develop or improve the safety, utility, effectiveness, efficiency, and usability of systems that include computers" (Diaper, 1989, p. 3)
- *Usability* and *HCI* are more or less separate disciplines that form *user experience* (Sherman & Quesenbery, 2005).
- Some take a holistic view of *usability* and *accessibility* by saying that “a universally accessible web site is one that is both usable and accessible” (Ivory & Chevalier, 2002, p. 2).

To ensure the consistent use of terms within this study, the term *usability* is considered an area within *HCI* as suggested by Diaper (1989). Additionally, *usability* (ISO, 1998) is distinguished from *accessibility* (ISO, 2003) as noted in the (ISO, 1998) definition provided above.

Web Usability

Web site user interfaces have unique features and demands which lead to the need for special attention and thus constitute a line of distinctive research and development

focused on Web usability within the broader area of *usability: Web usability*. Nielsen (2000a) has articulated the unique demands for Web design by stating that in traditional software environments, the designers hold the power of deciding product quality; in Web environments, the users share the power by virtue of the affordances of Web browsing software that allow users to change the look and feel of a Web site.

Usability can be viewed as more important for Web design than for traditional software design and requires extra attention. Why? Users of Web sites demand instant gratification. They usually refuse to waste their time with a low-quality Web site (a site that is confusing, slow, or doesn't satisfy their needs). As noted above, users, not just the designers, decide the quality of Web sites to a great extent. This degree of user empowerment with respect to Web sites should be a major concern for developers (Nielsen, 2000a). The developers of these sites should recognize that users are able to judge the usability quality of a Web site and quickly skip the site if they think the site is not easy to use. Nielsen (2000a) and other experts have called for more attention to usability issues in Web design.

Web Usability Evaluation

Web usability evaluation is a process that entails many methodologies for measuring the usability aspects of a Web site's user interface and identifying specific problems (Dix, 2003; Ivory, 2001; Nielsen, 1993). These methodologies allow collecting, analyzing, and critiquing usability evaluation data. Although usability evaluation employs many methods shared by both practitioners and academic researchers, the level of scholarly rigor of the methods being applied in usability testing varies greatly. For instance, a company may conduct a "practical" survey which consists

of questions generated by evaluators based on their anecdotal knowledge. In academic settings, a survey typically refers to a survey study which is guided by carefully-crafted research questions, organized by a sampling design, and yielding data that are analyzed using statistical methods. Most academic survey research is considered to be more valid and reliable than the informal survey research conducted by practitioners. In short, usability evaluation may generate findings that have various levels of scholarly rigor based on how the methods are implemented. The degree of scholarly rigor of a usability evaluation should be based on the objectives of that particular evaluation.

The following sections describe three major categories of usability evaluation (Reeves, Apedoe, & Woo, 2004): *usability testing*, *usability inquiry*, and *usability inspection*.

Usability Testing

Usability testing of a Web site is a process wherein the *intended users* of the site perform a set of predetermined tasks on the site while the *users'* behaviors are being observed by *evaluators* and recorded (see Figure 2.2). While the users are performing the tasks, they are typically encouraged to *think-aloud*: For example, a user 1) verbalizes the task while he or she is performing it, *I am clicking the contact button to look up the phone number*; and 2) verbalizes any obstacles he or she encounters, *I am at the contact page but don't see a phone number*. The goal of Web usability testing is to determine a site's level of ease-of-use based on users' experience and/or the extent to which a user is successful in accomplishing what the user comes to the site to do, e.g., make a purchase or find information. Usability testing is considered the most fundamental and effective method for evaluating user interfaces (Nielsen, 1993; Shneiderman, 1998).

A usability test may be conducted in a formal usability lab equipped with computers and audio and video recording devices; such a lab typically has two rooms, one for the user, the other for the evaluator(s) (Hughes & Burke, 2001). Usability tests can also be conducted in real-life settings, e.g., a classroom, using portable devices (Reeves & Hedberg, 2003).

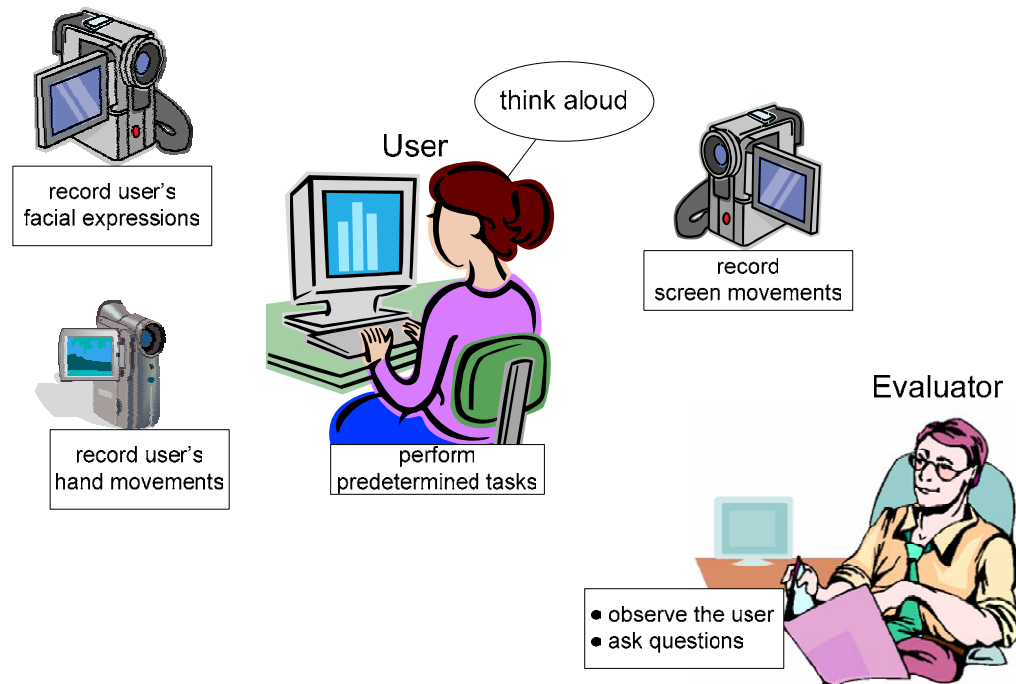


Figure 2.2. A usability test: User(s) perform predetermined tasks while being observed and recorded.

Ideally, all stakeholders (i.e., client, managers, designers, programmers, usability evaluators) of the target Web site should be involved in all parts of usability testing; however, practical limitations may force one person to take on multiple roles during the testing (Hughes & Burke, 2001; Reeves & Hedberg, 2003). An established “rule” exists which suggests that five users participating in usability testing is usually sufficient to identify the most critical usability problems (Nielsen, 2000b; Nielsen & Landauer, 1993; Virzi, 1992). However, others assert that having five users is far from enough for Web

usability testing (Molich *et al.*, 1999; Spool & Schroeder, 2001; Woolrych & Cockton, 2001). However, those objecting to just five users have not clearly established an alternative optimum number of users who should be participants in Web usability testing.

A formal usability test process consists of six major phases: planning, scenarios generation, walk-through, rehearsal, testing session (typically of one session per user), and resolution (Hughes & Burke, 2001) (see Figure 2.3).

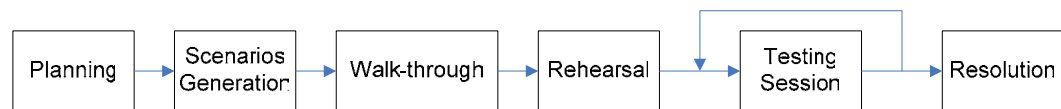


Figure 2.3. The six phases of a formal usability testing process.

First, the planning phase is conducted to identify users, define tasks and create a project plan. The planning phase produces a description of users' profiles as well as a list of tasks that will be performed by the users. During this phase, the evaluation team also lays out the plan of the entire project. As many stakeholders as feasible should be involved in the planning phase.

Second, during the scenarios generation phase, the task list is transformed into scenarios, i.e., high-level descriptions that will be given to the users during testing sessions. Selected member(s) of the evaluation team will create the scenarios.

Third, during a walkthrough, selected team member(s) perform the tasks based on the given scenarios. The purpose of a walk-through is to make sure that the scenarios are both correct and complete.

Fourth, during a rehearsal, selected users perform the tasks based on the given scenarios. The purpose of a rehearsal is to involve users to make sure that the scenarios are correct and complete.

Fifth, a *testing session* is where the actual usability testing occurs. During this phase, users perform tasks on the Web site based on the given scenarios. Users are typically asked to think aloud as they are performing the tasks. The evaluators observe and record the users' behaviors. If appropriate, the evaluators may also prompt the users or ask them to clarify their actions or thoughts. Normally, a usability test will have multiple test sessions, at least one session per user. Sometimes, a user may participate in multiple sessions. The data derived from each session are analyzed; the results of multiple sessions are combined.

Finally, the *resolution* phase summarizes, categorizes, and prioritizes the test results. During this phase, solutions to the usability weaknesses identified during the test are decided, an implementation plan is created to resolve the identified usability issues, and people are assigned the responsibilities to implement the solutions. All stakeholders should participate in the resolution phase to attain the best results. Figure 2.4 illustrates the six-phase usability testing process and the deliverables of each phase.

Usability testing is considered the most fundamental, complete, and effective approach for evaluating user interfaces; however, conducting a usability test can be very resource-intensive. In an ideal world where time, financial, human, and knowledge resources are unlimited, usability testing should be performed early, often, and extensively. However, real-world projects almost never have enough resources. In an environment where the evaluation team consists of one person, all aspects of the process should be performed to the best extent possible. Although the individual evaluator working alone must serve multiple roles, performing a usability test with one evaluator is better than not performing a test at all.

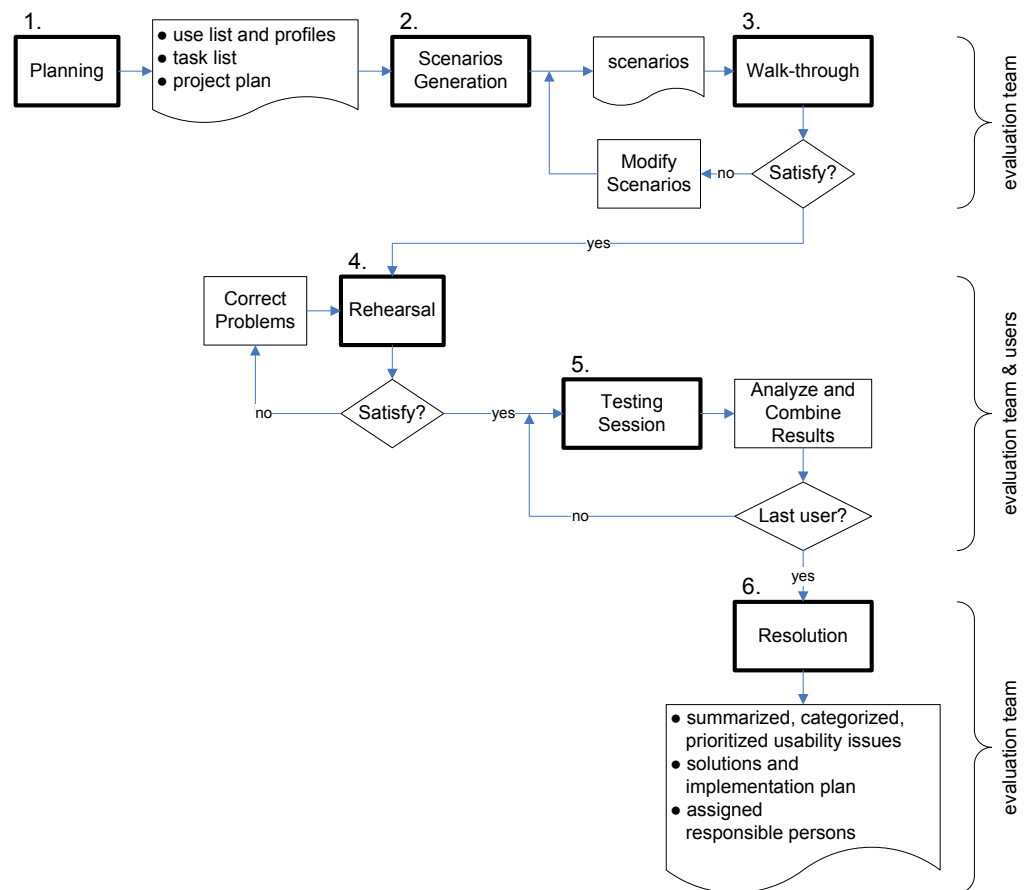


Figure 2.4. The usability testing process and the deliverables of each phase.

Usability Inquiry

Usability inquiry refers to approaches for collecting and analyzing users' subjective feedback (e.g., impressions, preferences, or opinions) about a Web site (Ivory, 2001; Reeves, Apedoe, & Woo, 2004). The goal of a usability inquiry is to evaluate users' overall experience, as opposed to the task-specific feedback gleaned from usability testing. There are two usability inquiry approaches:

- Supplemental approach—usability inquiry which shadows *usability testing* to achieve the dual goals of usability inquiry and usability testing (Ivory, 2001).

This model uses many research methods, mostly qualitative, to supplement usability testing methods.

- Variation approach— usability inquiry which utilizes usability testing methods, in addition to research methods (mostly qualitative), to achieve usability inquiry goals. However, in this model, the users explore a Web site freely, instead of performing pre-determined tasks (Bishop & Bruce, 2002; Hackos & Redish, 1998; Reeves, Apedoe, & Woo, 2004).

Usability inquiry methods include interviews, focus groups, field observation, and user logging (screen capturing).

Interviews. An interview is a discussion session between an interviewer (an evaluator) and an interviewee (a user). For purposes of usability evaluation, an interviewee is typically a person who has used the Web site under evaluation for a meaningful length of time (Reeves, Apedoe, & Woo, 2004). Three major forms of interview exist: *structured interview*, *semi-structured interview*, and *phenomenological interview*. The goal of a *structured interview* is to gather specific information about Web usability from a user (Hackos & Redish, 1998; Wilkinson, 2000). Interviewers typically ask prepared-questions that are intended to be numerous enough to “fill-up” the entire interview session (Wilkinson, 2000). A face-to-face survey allows the interviewer to ask follow-up questions and the user to ask clarifying questions. Structured interviews are more common in usability evaluations than unstructured ones (Reeves, Apedoe, & Woo, 2004). Although structured interviews are suitable for evaluating usability features that the evaluators intend to study, they risk the possibility that information not included in the list of questions may be overlooked. For instance, the interviewer’s question may be *Did you find all the items you expected to find in the pull-down list box?* Even if the

user's answer is *yes*, this question would not likely uncover the fact that the user was very frustrated in recognizing and locating the list box.

A *semi-structured interview* is a slight variation of a structured interview. Instead of using a strictly question-answer format, the interview is more conversational. An interviewer has interview objectives and uses guiding questions to steer the discussion with the interviewee in the directions related to usability issues.

The goal of a *phenomenological interview* is to understand the subjective *experience* of a user (Crotty, 1998; Seidman, 1998). In a phenomenological interview, interviewers are highly discouraged from using prepared-questions; instead, they are encouraged to start their interviews with short “experiential” questions (*Tell me about your experience in using . . .*), using subtle acknowledgements (*Aha, I see*) and a few prompts (*Tell me more about . . .*) (deMarrais, 1998). In a phenomenological interview, an interviewee is empowered to tell the stories of his or her own choosing (deMarrais, 1998). The role of an interviewer is to subtly make sure that the interviewee tells enough about each story for meaningful data analysis. Conducting a “proper” phenomenological interview is very difficult as an interviewer must be skillful in phenomenological techniques. Despite their challenges, phenomenological interviews are especially well-suited for understanding a user's frustrations in using a Web site.

Regardless of the style, interviews are typically audio recorded; sometimes they are even video recorded to capture an interviewee's body language. The recorded tapes are transcribed into text and analyzed. Transcribing audio to text is time-consuming and deceptively challenging; transcribing video to text, even more so.

Focus groups. A *focus group* shares similar goals and processes with an interview; an exception is that a focus group has one moderator and multiple participants (users) (Hackos & Redish, 1998; Wilkinson, 2000). Hackos and Redish (1998) have articulated some limitations of focus groups, “They don’t show behavior. They aren’t held in the users’ environment. They often include gatekeepers, not users. They may be dominated by a few individuals” (p. 146). Focus groups often follow the framework of structured or semi-structured interviews. Phenomenological focus groups (Hines, 2000; Mazzarol & Choo, 2003) do exist, but are seldom conducted.

Field observation. The goal of *field observation* is to uncover Web usability issues while the users are in their own environment. An evaluator observes user(s) performing their normal tasks while taking notes. In addition, the user may be video recorded, and the screen may also be captured on video as the user interacts with the site.

Surveys or Questionnaires. A *survey* (or *questionnaire*) is a set of questions about a Web site seeking users’ responses. Survey questions can be in many forms: true/false, multiple choice, short answer, and essay, and may be paper-based or Web-based. Creating a survey appears to be simple. However, creating a survey that addresses the right questions, ones which can yield meaningful results for improving Web usability, is very difficult (Fowler, 1995). All surveys are not created equally--one survey may be created by an evaluator based on his or her personal knowledge while another may be created empirically and organized by research design principles. Sound survey design principles demand that the survey should be based on clear objectives and context, guided by research questions, grounded in relevant theories and prior research studies, analyzed

by quantitative and qualitative research methods, and reported to ensure empirical soundness of the survey (Babbie, 1973; Fowler, 2002).

User logging. The goal of *user logging* is to automatically record movements on screen while a user is accessing a Web site using special software that works in tandem with the browser software. User logging is a powerful, detailed and economical usability inquiry technique. However, making sense of the results of user logging (screen capturing) for improving Web usability is difficult (Reeves, Apedoe, & Woo, 2004). Therefore, user logging should be used to supplement other usability evaluation methods such as usability testing or field observation.

Usability Inspection

Usability inspection is a process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria. Fundamentally, the goal of usability inspection is "to find usability problems in an existing interface design and then use these problems to make recommendations for improving the usability of an interface" (Ivory, 2001, p. 30). *Heuristic evaluation* (Nielsen, 1993; Nielsen & Molich, 1990) and *cognitive walkthrough* (Lewis, Polson, Wharton, & Rieman, 1990) are two common usability inspection methods. There are many other usability inspection methods as well. The following subsection highlights the following usability inspection methods: *heuristic evaluation*, *perspective-based inspection*, *feature inspection*, *guideline review*, *claims analysis*, *cognitive walkthrough*, and *pluralistic walkthrough*.

Heuristic evaluation. The goal of heuristic evaluation (Nielsen, 1993; Nielsen & Molich, 1990) is to identify possible usability problems of a Web site. Evaluators

independently explore a Web site and judge the overall site, individual pages and page elements based on a set of criteria. Heuristic evaluation criteria tend to be at a high-level (e.g., user control and freedom, or error prevention). When the evaluation is completed, a list of possible usability issues of the site is created based on the results of all evaluators. Heuristic evaluation is easy to learn and implement (Ivory, 2001) and is an economical approach to usability evaluation (Nielsen & Mack, 1994). Heuristic evaluation may also be implemented by following a traditional software inspection process, whereby evaluators from various areas-- developers, designers, documenters, trainers, technical support personnel, and usability experts--are assigned specific roles to follow to evaluate a Web site (Ivory, 2001).

Perspective-based inspection. The goal of *perspective-based inspection* (Zhang, Basili, & Shneiderman, 1999) is to identify possible usability problems of a Web site focusing on a specific viewpoint. The process of a perspective-based inspection is similar to the heuristic evaluation process with the exception that the evaluation criteria are divided into multiple “perspectives” or viewpoints (e.g., novice users, learners, frequent online shoppers). Each “perspective” is evaluated separately.

Feature Inspection. The goal of a *feature inspection* (Kahn & Prail, 1994) is to identify functional issues of a Web site based on the intended features of the site. Evaluators explore a Web site and analyze the availability and understandability of the site’s intended features. A feature list and set of accompanying scenarios are used for the evaluation. Documenters (e.g., technical writers) typically serve as evaluators of a feature inspection, but usability experts and users may be evaluators as well. A variation of feature inspection is the *consistency inspection* which has the goal of identifying

interface and functional inconsistency in sections of a Web site or multiple sites belonging to the same “family” (Ivory, 2001).

Guideline Review. The goal of a *guideline review* (Ivory, 2001) is to measure a Web site’s conformity to a set of established detailed guidelines. Evaluators explore a Web site and judge the site and its pages and elements based on a set of detailed guidelines, such as number of words in a page title or number of redundant links. At the end, the scores of the evaluator(s) are calculated and reported. A variation of guideline review is *standards inspection* which has the goal of measuring a Web site’s conformance to a set of established detailed organizational, industrial or governmental standards (Ivory, 2001).

Claims analysis. The goal of *claims analysis* (Carroll, 2000; Carroll & Rosson, 1992; Rosson & Carroll, 2002) is to identify the benefits and risks of usability features based on designers’ claims—the hypothesized effectiveness of the features (Keith, Blandford, Fields, & Theng, 2002; Rosson & Carroll, 2002). Evaluators list the possible positive and negative consequences of each user interface feature, as well as their trade-offs. For instance, using a map to represent the United States may be visually appealing to visual-oriented users. However, this feature may require longer access time than using a list which may possibly be preferred by frequent users who know where to go and want to get there quickly. In short, claims analysis is a methodical approach to evaluation usability feature trade-offs which leads to more informed usability design decisions.

Cognitive walkthrough. The goal of *cognitive walkthrough* (Blackmon, Polson, Kitajima, & Lewis, 2002; Lewis, Polson, Wharton, & Rieman, 1990) is to describe the usability successes and failures of Web access tasks. Evaluators explore a Web site

following a pre-determined set of tasks and record their experiences while assessing and recording the level of ease of each task. Cognitive walkthrough requires intensive documentation, and is considered hard to learn and implement. A variation of *cognitive walkthrough* is *cognitive jogthrough* (Rowley & Rhoades, 1992), whereby video recording and software logging are used to reduce the documentation intensiveness.

Pluralistic walkthrough. *Pluralistic walkthrough* (Bias, 1994) is another variation of cognitive walkthrough. In *pluralistic walkthrough*, multiple evaluators from multiple areas--usability experts, developers, users--perform the walkthrough as a group.

The Aspects of Web Usability Evaluation

Various aspects of a Web site or a Web page may affect the usability of the site or page. These aspects include the 1) page layout, 2) heading, titles, and labels, 3) navigation and links, 4) text appearance, and 5) graphics, images, and multimedia. This subsection highlights each of these aspects (Koyani, Bailey, & Nall, 2004).

Page layout. Page layouts involve structuring the Web pages for ease of comprehension. Elements of the page layout are the alignment of items on a page, the amount of white space on a page, the indication of the top or bottom of the page, the length of a line on the page, and the length of the page (Koyani, Bailey, & Nall, 2004).

Text appearance. There are several characteristics related to text appearance that help ensure that a Web site communicates effectively with users. These characteristics are the use of font type and size, the use of text color and contrast with backgrounds, the use of background color, the emphasis on important text, the use of attention-attracting features, such as animated text, and format consistency (Koyani, Bailey, & Nall, 2004).

Navigation and linking. Navigation refers to the method used to find information within a Web site. The usability of a Web site is determined by several features. The Web site must allow users to locate and link to destination pages and find and access information effectively. The usability is also determined by the length of the navigation page, the use of a site map, the use of feedback on the user's location within the site, the use of menu items, groupings of navigation elements, the use of descriptive tab labels on navigation elements, and the occurrence of dead-end pages (Koyani, Bailey, & Nall, 2004).

Linking refers to the Web element that, when clicked, will causes a new page to appear. Several factors that may affect the effectiveness of linking are the use of meaningful link labels, the consistency of clickability cues, the designation of visited links, and the use of text or graphics for links (Koyani, Bailey, & Nall, 2004).

Heading, titles, and labels. A Web page's heading, titles, and labels should facilitate both scanning and reading written material. Several factors that contribute to the effectiveness of these elements are the use of unique and descriptive headings and titles, the number of types and levels of headings, the appropriate use of HTML heading order, the descriptiveness of row and column headings of tables, and the highlighting of critical information (Koyani, Bailey, & Nall, 2004).

Graphics, images and multimedia. The appropriate use of graphics, images, and multimedia can add tremendous value to a Web site. Several factors that affect the usability of these elements are labeling images, the size and download speed of the graphics, images, and multimedia, the use of thumbnails, the use of images as the entire

background of a page, whether the images convey the intended message, and whether the elements resemble advertisements (Koyani, Bailey, & Nall, 2004).

In summary, Web usability encompasses challenges beyond software usability. *First*, a Web site is expected to function on multiple hardware platforms and Web browsers. *Second*, a Web site is not expected to be accompanied by an instructional manual, and Web users are not expected to be trained before using a site. *Third*, many Web sites are built by people who do not have an educational background in design and development. According to Clapsaddle (2004) “As more and more non-professionals attempt to design and publish their own websites, visitors accessing these websites grow increasingly frustrated due to the un-usability of these sites” (p. 3-4).

In response to the unique challenges of Web design, many resources include design guidelines devoted to improving Web usability design quality (Johnson, 2003; Koyani, Bailey, & Nall, 2004; Krug, 2000; Nielsen, 2000a; Nielsen & Tahir, 2002; Ratner, 2003; Sterne, 2002; Weinman & Karp, 2003). Designers are advised to apply these principles when designing Web sites, but following the principles is not like following a simple recipe. Since their application does not guarantee perfectly usable Web sites, the resulting sites must still be evaluated and systematically enhanced. Usability *testing*, *inquiry*, and *inspection* are three usability evaluation approaches to help evaluators and designers evaluate and improve Web sites systematically. Each of these methods has advantages and drawbacks, so evaluators must choose and mix various methods and apply them based on the specific context. Automated Web usability evaluation is another approach to facilitate or to supplement usability evaluation to

address the challenges of Web usability improvement. As described in Chapter 1, this dissertation focused on automated Web site usability evaluation.

Automated Web Site Usability Evaluation Methods and Tools

An automated Web site usability evaluation tool is a software program that analyzes Web site user interfaces to determine the usability quality. Some tools are fully automated, others are partially automated. Although most tools are labeled as *usability* evaluation tools, some, like WebXACT, emphasize *accessibility* (Watchfire, 2004), other tools, such as the program LIFT (UsableNet, 2004), analyze both usability and accessibility, and still others, such as the program WebTango (Ivory-Ndiaye, 2004b), minimize the distinctions between *usability* and *accessibility*.

Ivory (2003) synthesizes Web usability evaluation into five major approaches (see Table 2-1): *usability testing*, *inspection* (heuristic evaluation, cognitive walkthroughs), *inquiry*, *analytical modeling*, and *simulation*. The approaches may share one or more of these activities (see Table 2-2): *capture*, *analysis*, or *critique*. These activities may also be supported by one or more types of automated evaluation tools (see Table 2-3): *performance measurement*, *log file analysis*, *guideline review*, *textual analysis*, or *information-seeking simulation*. Ivory's (2003) synthesis of Web usability evaluation approaches, activities, and tools is illustrated in Figure 2.5.

Potential Benefits of Using Automated Evaluation Tools

The main motivation for automating Web site usability evaluations is to reduce the amount of usability testing performed manually.

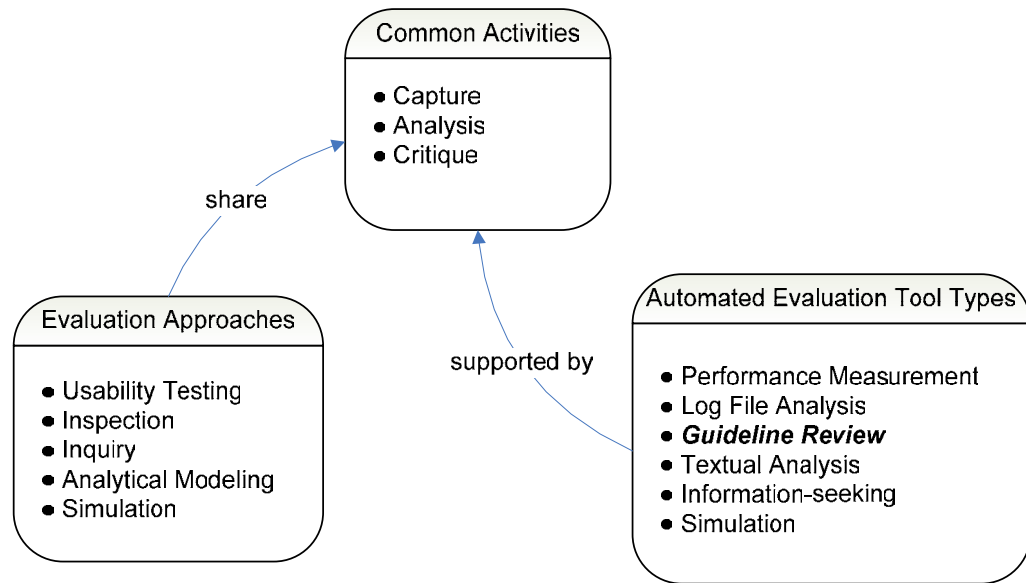


Figure 2.5. The dimensions of usability evaluation approaches, activities, and tools (Ivory, 2003). Automated **guideline review**, highlighted, was the focus of this dissertation which will be discussed in more detail in section Automated Guideline Review Tools.

Table 2-1

Usability Evaluation Approaches

Evaluation Approaches	Description
Usability Testing	“an evaluator observes participants interacting with an interface (i.e., completing tasks) to determine usability problems” (Ivory, 2003, p. 112).
Inspection	“an evaluator uses a set of criteria or heuristics to identify potential usability problems in an interface” (Ivory, 2003, p. 112).
Inquiry	“participants provide feedback on an interface via interviews, surveys, etc.” (Ivory, 2003, p. 112).
Analytical Modeling	“an evaluator employs user and interface models to generate usability predictions” (Ivory, 2003, p. 112)
Simulation	“an evaluator employs user and interface models to mimic a user interacting with an interface and report the results of this interaction (e.g., simulated activities, error, and other quantitative measures)” (Ivory, 2003, p. 112)

Table 2-2

Usability Evaluation Activities Shared by Evaluation Approaches

Evaluation Activities	Description
Capture	“collecting usability data, such as task completion time, errors, guideline violations,, and subjective ratings” (Ivory, 2003, p. 112).
Analysis	“interpreting usability data to identify problems” (Ivory, 2003, p. 112).
Critique	“suggesting solutions or improvements to mitigate problems” (Ivory, 2003, p. 112).

Table 2-3

Types of Evaluation Tools Supporting Evaluation Activities

Types of Evaluation Tools	Description
Performance Measurement	“monitor the consistency, availability, and performance of a web server or stress the server to determine the amount of traffic that it can accommodate” (Ivory, 2003, p. 113).
Log File Analysis	“identify potential problems in usage patterns (e.g., pages that are not being visited, broken links, or server errors)” (Ivory, 2003, p. 114).
Guideline Review	“detect and flag a web page’s or site’s deviation from design criteria” (Ivory, 2003, p. 114).
Textual Analysis	“identify potential navigation problems on web pages that are attributable to confusing headings or link text” (Ivory, 2003, p. 115)
Information-seeking Simulation	“mimic the browsing process of users and output computed measures and actions (e.g., navigation time and navigation paths)” (Ivory, 2003, p. 115)

Automated evaluation serves a similar purpose as heuristic evaluation and has many potential benefits (Brajnik, 2004; Clapsaddle, 2004; Ivory, 2003; Laskowski, Landay, & Lister, 2002; Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004):

- Automated evaluation reduces the cost of usability evaluation.
- It allows more in-depth analysis in certain aspects.
- It detects errors exhaustively.
- It reduces the manpower and finances needed for evaluations. Studies have suggested that a Web site usability test requires many more than five users (Molich et al., 1999; Spool & Schroeder, 2001). Manual usability evaluation

can be expensive and time consuming, especially evaluations that involve more than five participants.

- It increases consistency in error detection and feature evaluation.
- It increases consistency and repeatability of evaluation process.
- It can be performed during Web site design and development or during post-production.
- It can be used by novice Web builders, as well as professional designers.
- It reduces the need to have usability experts to evaluate Web sites.

Limitations of Using Automated Evaluation Tools

The proponents of automated Web site usability evaluation tools consistently emphasize that automated evaluation tools are considered supplements of, not replacements for, manual evaluation methods or techniques such as usability testing or heuristic evaluation (Brajník, 2004; Clapsaddle, 2004; Ivory, 2003; Laskowski, Landay, & Lister, 2002; Vanderdonck, Beirekdar, & Noirhomme-Fraiture, 2004). An automated Web site usability evaluation tool is like “a kind of ‘quality checker’ tool, similar in analogy to a spell checker in a word processor” (Ivory & Hearst, 2002, p. 367). Even a spell-checked document should be proofread by a human reader.

Some shortcomings of automated usability evaluation tools have been identified. *First*, automated evaluation tools do not offer a complete solution to usability evaluation. *Second*, “subjective measures such as user satisfaction are unlikely to be predictable by automated methods” (Ivory, 2003, p. 113). *Third*, a tool may misinterpret Web sites and produce false-positive results or false-negative results. *Finally*, Web site builders might over-rely on automated tools.

Automated Guideline Review Tools

An automated *guideline review* tool, one of several types of evaluation tools (see Figure 2.5), is a usability evaluation tool that checks the level of conformity of a Web site to a set of established Web interface guidelines (Ivory, 2003; Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004). A *guideline review* tool may involve one or more evaluation activities: *capture*, *analysis*, or *critique* (Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004) (see Figure 2.5 and Table 2-1). The following are samples of *usability* or *accessibility* evaluation guidelines: Section 508 (Center for Information Technology Accommodation, 2002), W3C guidelines (W3C, 2004), and other expert guidelines (Johnson, 2003; Koyani, Bailey, & Nall, 2004; Krug, 2000; Nielsen, 2000a; Nielsen & Tahir, 2002; Sterne, 2002; Weinman & Karp, 2003). Web design guidelines have been used manually by designers and developers, not necessary by usability experts, to improve Web sites' usability quality (Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004). The use of automated *guideline review* tools may help overcome some shortcomings of non-automated guideline-based usability evaluation such as the subjective interpretation of guidelines by designers (Ivory & Hearst, 2001). With proper design architecture, an automated guideline review tool allows for the use of new guidelines, custom guidelines, a subset of guidelines, or differing levels of priority. Interestingly, Vanderdonckt et al. (2004) found that users reported that using automated tools allowed them to learn about usability design. As described in Appendix A: Automated Guideline Review Usability Evaluation Tools, Ivory (2003) has identified 13 automated *guideline review* usability evaluation tools (see Table A-1). The next subsection elaborates on one of these automated *guideline review* tools: WebTango.

WebTango

WebTango is a *guideline review* automated Web interface evaluation approach intended to help novice Web site builders to create high quality Web interface designs or at least steer them away from bad designs (Ivory & Hearst, 2002). WebTango was originally designed for *usability evaluation* and *performance evaluation of informational sites*. WebTango evaluates a Web site based on 157 empirically evaluated Web site usability measures in nine categories: Graphic Elements, Graphic Formatting, Link Elements, Link Formatting, Page Formatting, Page Performance, Site Architecture, Text Elements, and Text Formatting (see Appendix B: The WebTango Web Interfaces Usability Measurements for more details). WebTango is intended to help users to operationalize Web design guidelines or heuristic evaluation.

WebTango consists of three major components: 1) *WebTango Analysis Tool* (Ivory-Ndiaye, 2004b), for evaluating the site's usability, 2) *TangoViewer* (Ivory-Ndiaye, 2004a) for displaying the analysis results produced by the *WebTango Analysis Tool*, and 3) *WebTango Models* (or statistical profiles), for predicting quality Web pages and sites. *WebTango Analysis and TangoViewer Tools*.

WebTango Analysis tool is a Web-based application that accepts a Web address, analyzes the usability measures of the site, and generates a collection of compressed text files containing coded quantitative measures and qualitative predictions. The analysis results of WebTango Analysis are not intended to be viewed by users. Instead, TangoViewer is a Java-based application that translates the coded results of WebTango Analysis and displays the results in a human readable format. Figure 2.6, Figure C.1, and Figure C.2 in Appendix C show sample screens of WebTangle Analysis and

TangoViewer tools. Figure 2.7 illustrates, conceptually, an overview of the process of evaluating a Web site using WebTango.

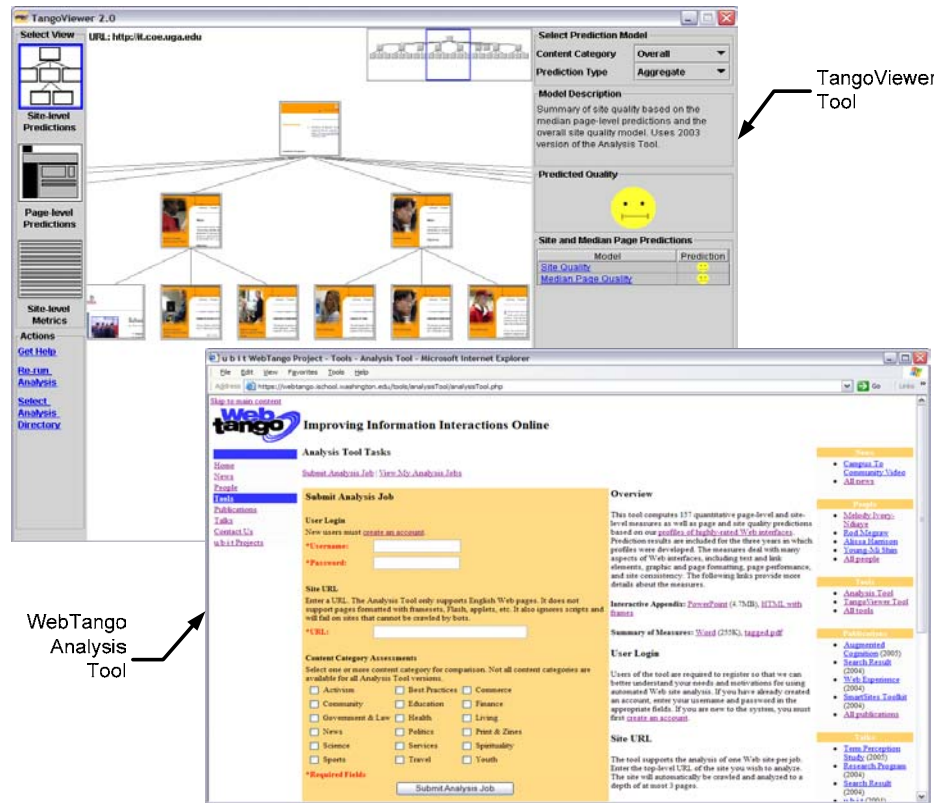


Figure 2.6. WebTango Analysis tool and TangoViewer tool screen samples.

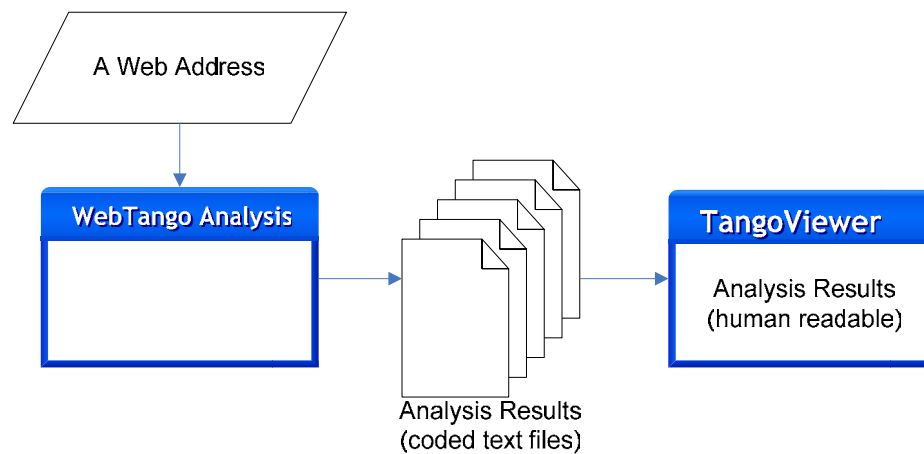


Figure 2.7. WebTango Evaluation process overview.

Actually, several steps are necessary to use these tools. In order to use WebTango to analyze a Web site, a user first has to 1) set up an account, and 2) download and install TangoViewer on a local computer. Both of these tasks start at <http://ubit.ischool.washington.edu/pages/tools.php> (see Figure C.3 in Appendix C for detailed steps). The following steps are required to analyze a Web site using WebTango:

1. Submit a URL at
<https://webtango.ischool.washington.edu/tools/analysisTool/analysisTool.php>.
2. Receive an e-mail from WebTango containing the URL of the compressed (tar) file of the analysis results.
3. Download the compressed (tar) result file and store on a local computer.
4. Decompress the tar file into resulting a coded (computer readable) analysis result file using a file compression/decompression program (e.g., WinZip).
5. Use TangoView to open the coded analysis result file.
6. View the analysis results in TangoViewer.

Figure C.4 in Appendix C illustrates the steps taken by a user to analyze a Web site using WebTango.

Users interact with WebTango Analysis and TangoViewer tools directly. However, an important component of WebTango, the models, is transparent to the users of WebTango. The next subsection describes WebTango models that serve as the conceptual engine of WebTango.

WebTango Models

Statistical Profiles of Highly-rated Web Sites (Ivory, 2001, 2003; Ivory & Hearst, 2002) are sets of decision tree rules for predicting usability quality of Web pages. (When

discussing WebTango, the terms *statistical profile* and *model* are used interchangeably.)

WebTango models are the algorithms used by WebTango Analysis tool to evaluate or predict the usability quality (good, average, poor) of a Web site or Web page. According to Ivory and Hearst (2002), “What distinguishes our work [WebTango] from most others is that this tool is based on empirically-derived measures computed over thousands of web pages” (p. 367). Figure 2.8 illustrates one example of one of the WebTango models.

if ((Italicized Body Word Count is missing OR (Italicized Body Word Count ≤ 2.5)) AND (Minimum Font Size is missing OR (Minimum Font Size ≤ 9.5)) AND (Graphic Ad Count is not missing AND (Graphic Ad Count > 2.5)))
Class = Good

This rule classifies pages as good pages if they have: two or fewer italicized body text words; use a font size of 9pt or less for some text; and more than two graphical ads.

if ((Italicized Body Word Count is missing OR (Italicized Body Word Count ≤ 2.5)) AND (Minimum Font Size is missing OR (Minimum Font Size ≤ 9.5)) AND (Graphic Ad Count is missing OR (Graphic Ad Count ≤ 2.5)) AND (Exclaimed Body Word Count is missing OR (Exclaimed Body Word Count ≤ 12.5)) AND (Exclaimed Body Word Count is not missing AND (Exclaimed Body Word Count > 11.5)) AND (Bobby Priority 2 Errors is missing OR (Bobby Priority 2 Errors ≤ 5.5)) AND (Meta Tag Word Count is missing OR (Meta Tag Word Count ≤ 66)) AND (Emphasized Body Word Count is missing OR (Emphasized Body Word Count ≤ 174.5)) AND (Bad Panel Color Combinations is missing OR (Bad Panel Color Combinations ≤ 2.5)))
Class = Average

This rule classifies pages as average pages if they have: two or fewer italicized body text words; use a minimum font size of 9pt or less for some text; two or fewer graphical ads; twelve exclaimed body words (i.e., body text followed by exclamation points); five or fewer Bobby priority 2 errors; 66 or fewer meta tag words; 174 or fewer emphasized body words (i.e., body text that is colored, bolded, italicized, etc.); and less than two bad panel color combinations.

if ((Italicized Body Word Count is not missing AND (Italicized Body Word Count > 2.5)))
Class = Poor

This rule classifies pages as poor pages if they have more than two italicized body text words.

Figure 2.8. A sample WebTango page-level prediction model (good, average, and poor) (Ivory, 2001, p. 156).

WebTango models, consisting of Web design guidelines, serve as the evaluation engine of the WebTango Analysis tool. WebTango models are used to predict the quality of Web sites and pages; their initial results were validated by comparing their predictions with results from the 2000 Webby Awards (Webby, 2005).

The WebTango models were developed by an iterative process (Ivory & Hearst, 2002; Ivory, Sinha, & Hearst, 2000, 2001). The latest and most extensive version has the highest number of measures (157), uses the highest number of Web sites (5346 pages from 639 sites), and produces the most accurate predictions (Ivory & Hearst, 2002). Figure 2.9 and Figure 2.10 illustrate WebTango *page-level* and *site-level* models' development process and the areas of results.

WebTango's Limitations

Similar to other automated usability evaluation tools, WebTango is intended to supplement other usability evaluation approaches such as usability testing, inspection, inquiry, analytical modeling, and simulation, not replace them. WebTango is "a kind of 'quality checker' tool, similar in analogy to a spell checker in a word processor" (Ivory & Hearst, 2002, p. 367).

The following list highlights WebTango's known limitations. *First*, WebTango only processes static Web pages and server-based scripting pages like *php*, and *jsp*, and does not support framesets or Web elements embedded in Web pages (e.g., Flash, ActiveX, Java Script, and Java Applets). *Second*, WebTango evaluates Web sites up to a depth of only three pages. *Third*, WebTango only supports English Web pages. *Fourth*, users are currently required to take many steps in order analyze a Web site using WebTango (see Figure C.3 and Figure C.4 in Appendix C). Although many of these

steps may be technically necessary, it would be better if they were transparent to the users. Ideally, WebTango Analysis Tool and TangoViewer Tool should be integrated into one tool.

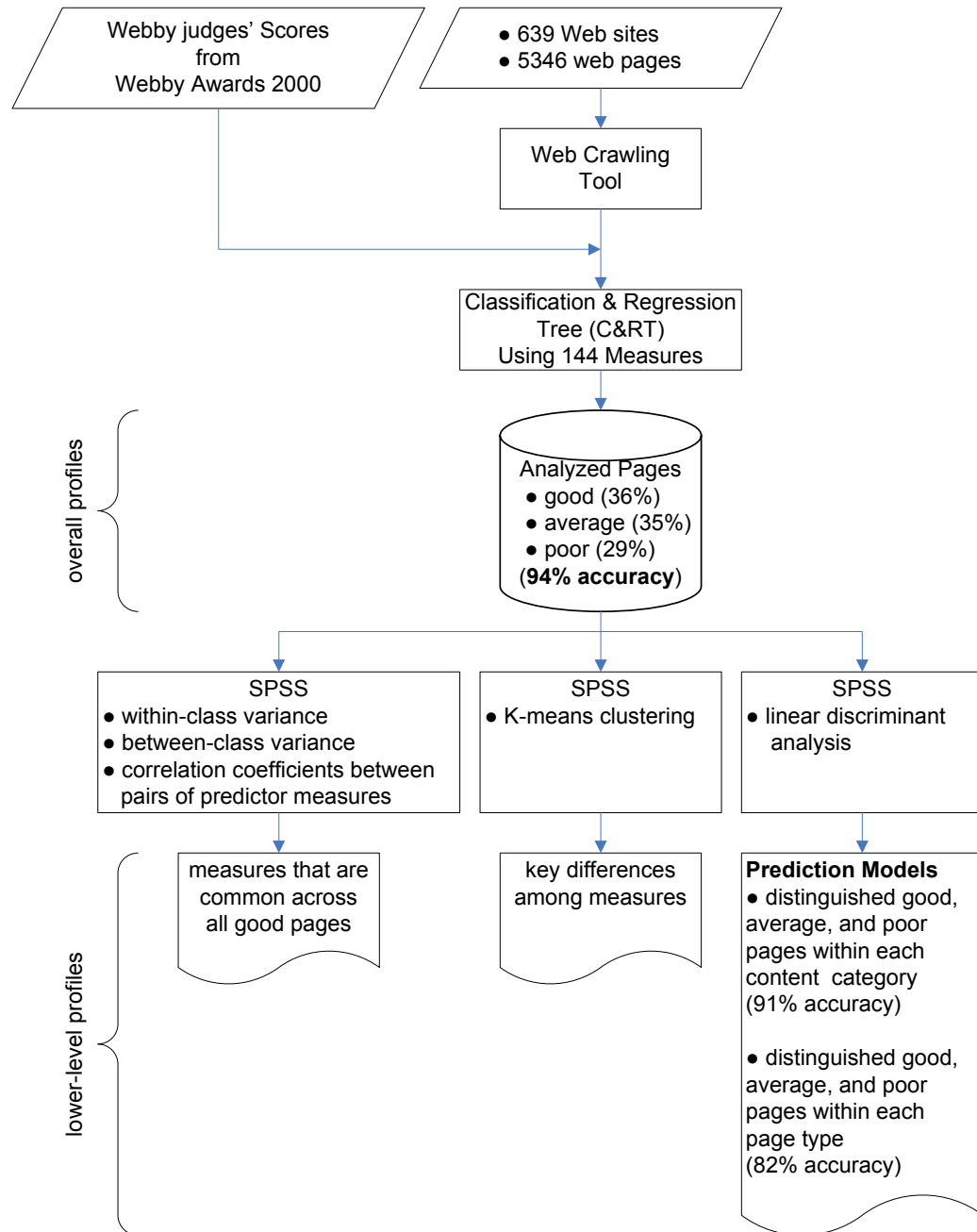


Figure 2.9. WebTango page-level measurement model development process.

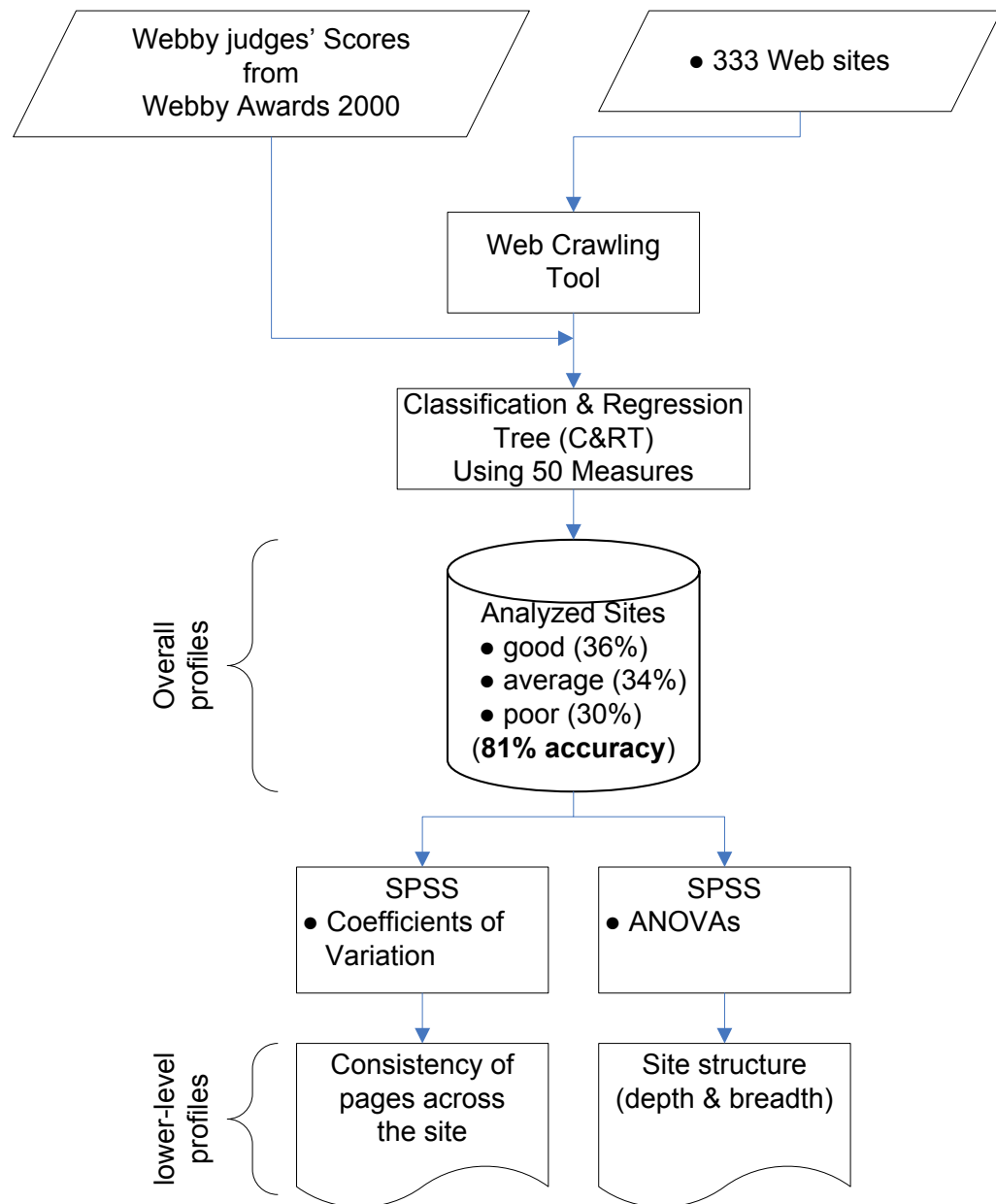


Figure 2.10. WebTango site-level measurement model development process.

In summary, WebTango is a guideline review-automated-Web-interface-evaluation approach that has a result-viewing tool and an analysis tool, with empirically validated evaluation guideline models that serve as the engine of that analysis tool. The development process of WebTango is very well documented. WebTango is not only

grounded in empirical models, but also appears to be the tool that has been most widely used in research. More extensive material about WebTango can be found in the following literature: *Automated Web Site Evaluation: Researcher's and Practitioner's Perspectives* (Ivory, 2003), *Statistical Profiles of Highly-rated Web Sites* (Ivory & Hearst, 2002), and *An Empirical Foundation for Automated Web Interface Evaluation* (Ivory, 2001).

WebTango Related Research Studies

Linking Web Interface Profiles to Usability

In the process of building WebTango evaluation models, Ivory (2001) conducted a study to determine the relationship between Webby 2000 judges' scores of Web site usability (Webby, 2005) and ratings assigned by human participants based on the Website Analysis and MeasureMent Inventory (WAMMI) (Kirakowski, Claridge, & Whitehand, 1998). The study addressed the question of whether Webby judges' scores are consistent with usability ratings by participants. Thirty participants, mostly University of California at Berkeley undergraduates, evaluated 57 Webby awarded Web sites based on WAMMI. The participants visually rated these Web sites, as well as performed information-seeking tasks as part of the evaluation. Ivory (2001) found Webby judges' scores were "mostly consistent" (p. 199) with participants' ratings. However, the findings are not statistically conclusive. Ivory (2001)'s study suggests that the WebTango model "reflect[s] usability to some degree" (p. 199). Figure 2.11 provides an overview of this *Linking Web Interface Profiles to Usability* study.

Ivory (2001)'s study provided encouraging, but inconclusive, results for empirically supporting the validity of WebTango models. There were at least two major

possible shortcomings of this study: 1) The participants mostly consisted of undergraduate (and few graduate) students. 2) The Web sites evaluated were not identical to the sites evaluated by Webby judges because the study was conducted at least six months after Webby Awards 2000.

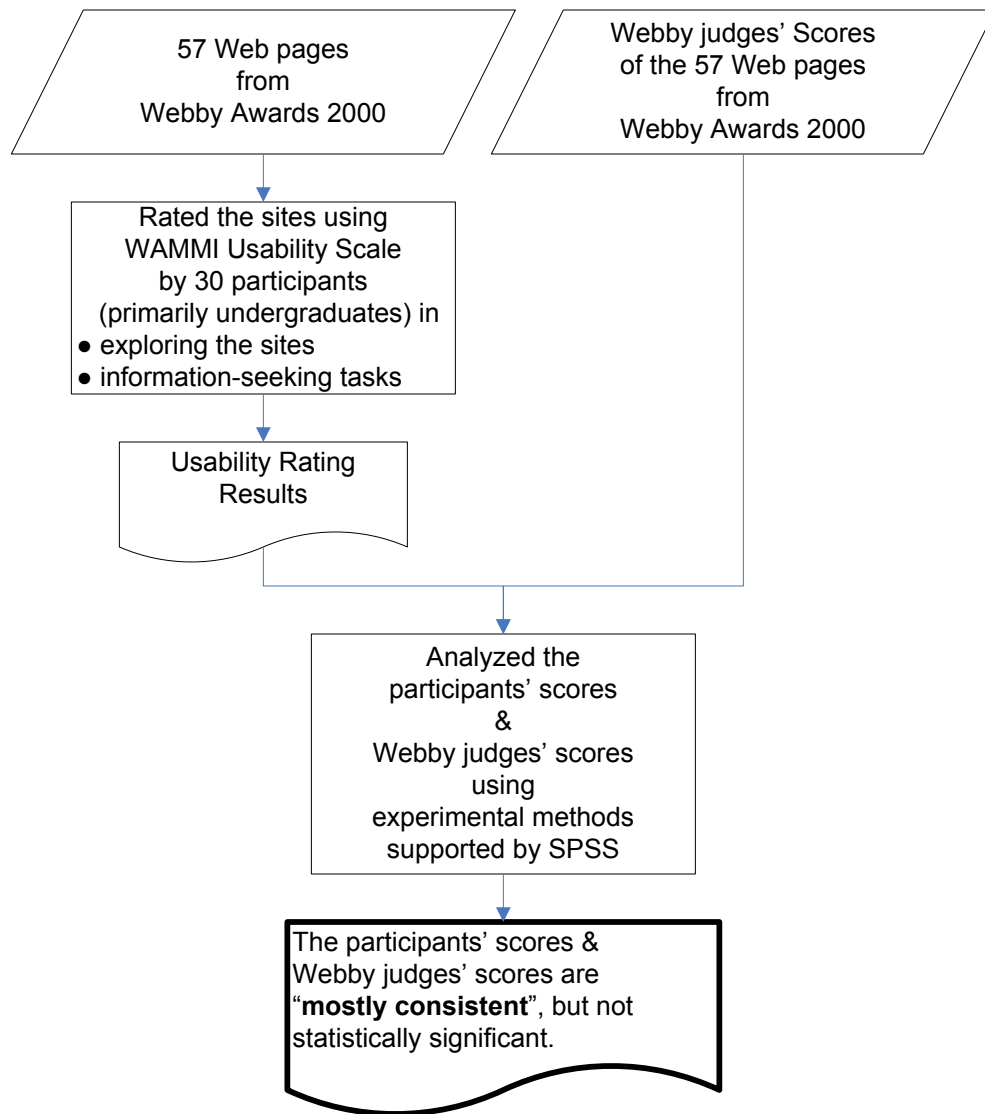


Figure 2.11. An overview of Linking Web Interface Profiles to Usability study.

Evaluating the Web Interface Profiles

Ivory (2001) also conducted a study, *Evaluating the Web Interface Profiles*, to determine whether changes made to Web pages based on two of the WebTango models, *overall page quality model* and *good page cluster model*, actually improved design quality. The hypothesis of this particular study in Ivory (2001) was that the “pages [or sites] modified based on the [WebTango] overall page quality and the good page cluster models are of a higher quality than the original pages [or sites]” (pp. 238-239). Thirteen participants rated the *original Web pages* and *modified Web pages* (thirty-one Web pages from five Web sites) in two different ways: 1) They selected the pages that they considered high-quality, and 2) they rated each site on a 5-point Likert scale. The results indicated that participants to a statistically significantly degree preferred the modified pages and rated the modified sites higher than the original ones. Figure 2.12 illustrates the overall process and the results of this study.

Although the results of this study were positive, there may have been possible shortcomings. For starters, *the study only tested the perceived usability condition*. Ivory (2001) indicated that because only a subset of Web pages belonging to the five Web sites were modified, the participants only evaluated the sites visually; *information-seeking tasks* were not tested.

Another possible problem with Ivory’s study was that *diverse Web site categories were tested*. Five sites selected from various Yahoo categories (education, health, community, finance, and living) were used in the study. Given that the category of a Web site appears to have an impact on usability evaluation results, one study that tested five sites from five different categories may not have provided the most accurate results.

A final possible problem with the study was that, *WebTango models only partially influenced the study results*. The students modified the Web pages mainly guided by WebTango's *overall page quality model* and *good page cluster model*, but Web pages were also modified partly based on students' intuition. However, Ivory (2001) stated that "The students had little or no training in Web design and had very little experience with building Web sites" (p. 220); moreover, "students had to rely on their own intuition in cases w[h]ere design changes were not as straightforward" (p. 220). The extent to which the students modified the pages based on their intuition was not stated. Of course, the use of an automated evaluation tool alone is not sufficient to improve Web site quality. In addition, not all measures were used to modify the pages; for example, the content of the pages, with minor exceptions, remained the same.

A Study of Automated Web Site Evaluation Tools

Ivory and Chevalier (2002) conducted an empirical study aimed at determining whether using automated evaluation tools improves Web sites from both the designer's and the user's perspectives. The hypothesis of Ivory and Chevalier (2002) was that "using an automated evaluation tool to guide site modifications would enable designers to identify more site problems, to correct more of these problems, and to consequently better improve the usability and accessibility of the site subsections, at least more so than not using automated evaluation tools" (p. 5). Ivory and Chevalier's study consisted of three sub-studies. *First*, Ivory and Chevalier surveyed 169 Web design professionals concerning their work practices related to *universal accessibility*, *usability* and the use of automated usability evaluation tools.

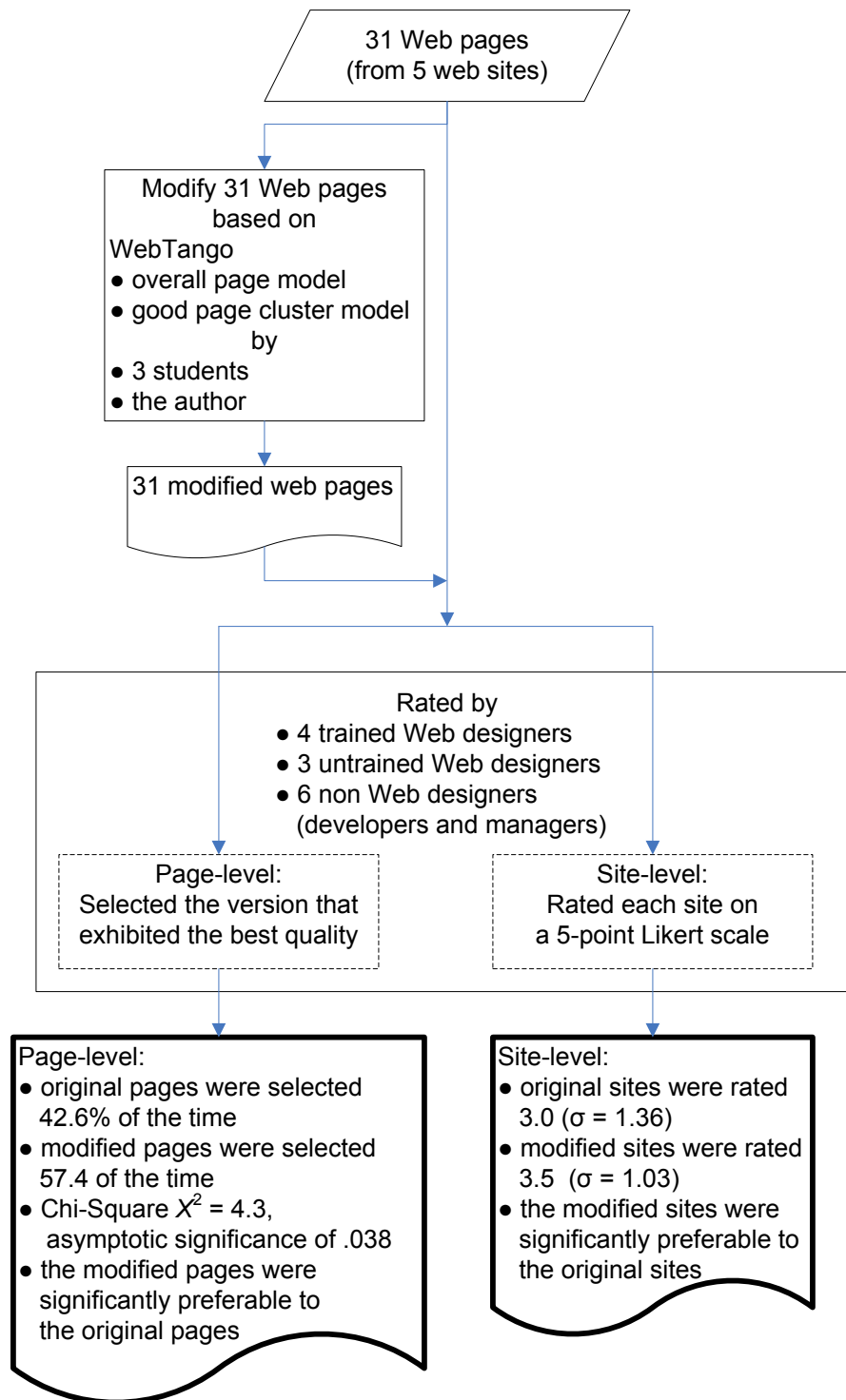


Figure 2.12. An overview of the Evaluating the Web Interface Profiles study.

Next, they studied how experienced Web designers used three *guideline review* automated usability evaluation tools (WatchFire bobby, WEC HTML Validator, and UsablerNte LIFT) to modify five Web sites. Finally, the researchers conducted a usability study wherein 22 users “with and without visual, physical, and learning impairments completed information-seeking tasks on the original and modified sites” (Ivory & Chevalier, 2002, p. 1). Figure 2.13 illustrates the overall process and the results of Ivory and Chevalier’s study.

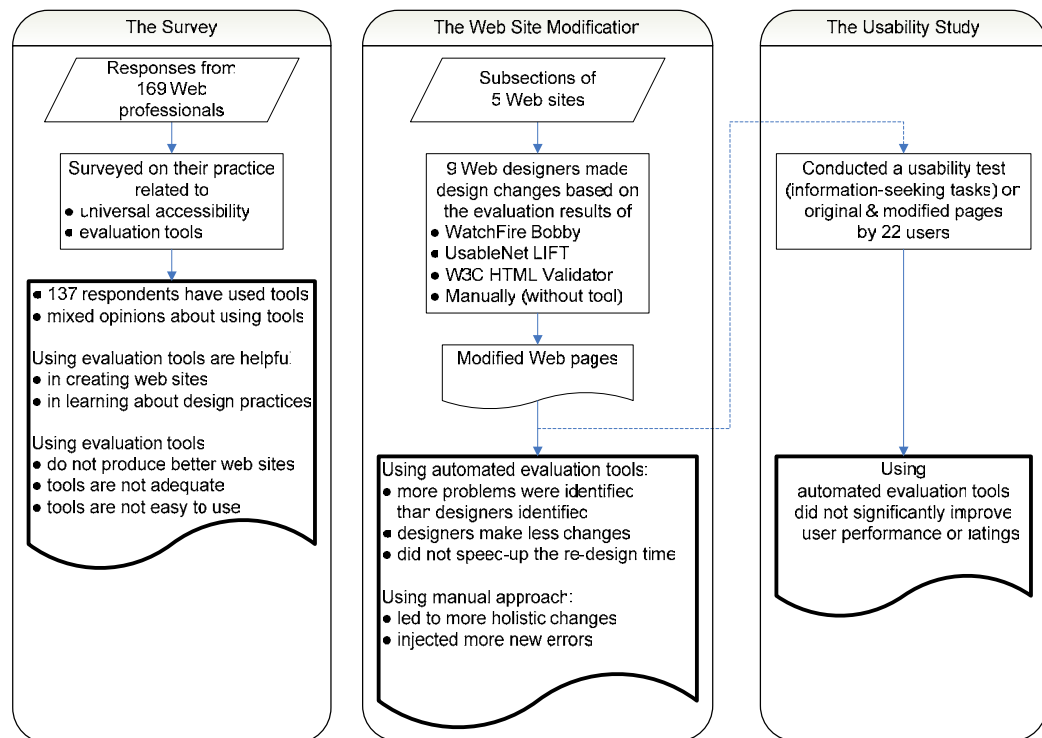


Figure 2.13. An overview of A Study of Automated Web Site Evaluation Tools.

The researchers were surprised by the main findings of their study (Ivory & Chevalier, 2002): Modifying Web sites using automated evaluation tools did not improve user performance or ratings to a statistically significant degree. Several possible reasons exist for Ivory and Chevalier (2002)’s disappointing results.

Combining both usability and accessibility in one hypothesis may be too broad.

Although Ivory and Chevalier (2002)'s holistic philosophy, "a universally accessible web site is one that is both usable and accessible" (p. 2), is sensible, in the context of one specific research study, combining *usability* and *accessibility* may lead to insensible results.

The majority of the usability test participants had various forms of disabilities.

People with disabilities have special needs that are not typically addressed by usability guidelines. The 22 participants worked in very different conditions with diverse usability and accessibility needs: "Their ages ranged from 18 to 55,.... Thirteen participants had a visual (6), hearing (1), moderate mobility (3), learning (1), physical (5), or other (1) disability. Three participants actually had multiple disabilities" (Ivory & Chevalier, 2002, pp. 11-12). Although the study covered both *usability* and *accessibility* problems, changes made on usability and accessibility related problems were not separately reported. Therefore, the modifications made may have been related to the usability needs but not to the accessibility needs of the participants. Additionally, since the participants had very diverse forms of disabilities that needed to be addressed differently, their accessibility needs (e.g., people with visual, hearing, or learning disabilities) must be addressed separately as well.

The number and mixture of participants may not be appropriate for a Web

interface usability study. The 22 participants ranged from 18 to 45 years of age and had many forms of disabilities: six visual, one hearing, three moderate mobility, one learning, five physical, one other, and three multiple disabilities. The wide variety of user ages and abilities may have skewed the results.

The time allocated to modify the Web sites may be too short or unrealistic. Each of the nine Web designers modified five Web sites. Twenty minutes were allocated to modify four sites, and forty minutes to modify the fifth Web site. The time pressure factor may have led to unrealistic findings.

The use of usability evaluation tools is helpful but is not sufficient to improve Web site quality. Experts in automated usability evaluation tools have consistently emphasized that evaluation tools should be considered as supplements to other usability evaluation approaches; using evaluation tools without other approaches is not sufficient to improve Web site quality. The results of Ivory and Chevalier (2002)'s study may confirm the premise that using an automated evaluation tool alone is not sufficient to improve Web site quality.

In summary, the surprising results of Ivory and Chevalier (2002)'s study may have been impacted by the following issues: 1) In addressing usability, associability, and the effectiveness of evaluation tools, their aim was too broad, 2) the participants of the usability test were too diverse, 3) the time allocated to modify the sites was too brief, and 4) the automated tool was not used as a supplement to other usability evaluation approaches as recommended in the literature. A synthesis of multiple but separate research studies, each addressing a specific research question, may be a more appropriate way to study the question intended by Ivory and Chevalier (2002), namely, whether “using an automated evaluation tool to guide site modifications would enable designers to identify more site problems, to correct more of these problems, and to consequently better improve the usability and accessibility of the site subsections, at least more so than not using automated evaluation tools” (p. 5).

Measuring Usability: Categorically Modeling Successful Websites Using Established Metrics

Clapsaddle (2004) developed a predictive, categorized model of an ideal Web site, based on an established set of empirical usability metrics in order to enable the novice to design better Web sites. Clapsaddle's approach to model development was similar to Ivory (2001)'s approach in developing the WebTango model. Clapsaddle's model is basically an extension of the WebTango model. Her study produced two sets of deliverables:

- a list of metrics ranked in the order of their importance when classifying Web pages by usability rating (see Figure 2.16 for a partial list)
- categorical models, and associated rule sets for developing successful Web pages (see Figure 2.15 for a sample model).

The following list (and Figure 2.14) highlights the methodology of this study:

- One hundred-twenty Web sites (2631 pages) were used as the dataset for this study.
- WebTango (Ivory-Ndiaye, 2004b) was used to generate 157 empirical Web interface usability measures. Doctor HTML (Doctor HTML, 2003) and Linklint (Bowlin, 2001) were used to generate measures that are not supported by WebTango: link validity, number of cookies, and browser compatibility.
- The research visually determined the quality of 120 homepage layouts (i.e., the positions of Web components, such as buttons, graphics, Flash components, etc.). Each homepage was assigned a grade of good, average or poor. This visual detracting was based on various Web page layout guidelines

(Bernard, 2000, 2001, 2002; Bernard, Baker, & Fernandez, 2002; Bernard & Hull, 2002; Bernard, Liao, & Mills, 2001; Bernard, Mills, Frank, & McKown, 2001; Markham & Hall, 2003).

- The researcher manually categorized the 120 Web sites into four categories: e-commerce, informational, interactive, and portal.
- WEKA (Frank et al., 2005), a data mining software program, was used to “mine for trends that may be useful in predicting the usability of a web site” (Clapsaddle, 2004, p. 69).

Clapsaddle (2004) employed both heuristic and empirical methodologies in her study. The following list summarizes the results and implications of this study:

- The study created a ranked list of Web site usability metrics of all four sites categories (i.e., e-commerce, informational, interactive, and portal). Figure 2.16 contains the top 25 measures of e-commerce. The metrics implicate that the predictability of Web site usability is different for Web sites that belong to different categories (i.e., e-commerce, informational, interactive, and portal)
- This study created a set of decision tree classifier models; each model contains a rule set for a “good” Web site. Figure 2.15 illustrates a rule set for “good” e-commerce Web sites. These models may be used to “derive a conformance index, a measure of the extent to which a Web site design has followed the established guidelines” (Clapsaddle, 2004, p. 107), for each of the Web site categories.
- This study found that non-measures such as link validity, number of cookies, and browser compatibility, are important measures of Web site usability and

should be included when using automated Web usability evaluation tools.

Additional research is necessary to ascertain the relative importance of these measures to a specific Web category.

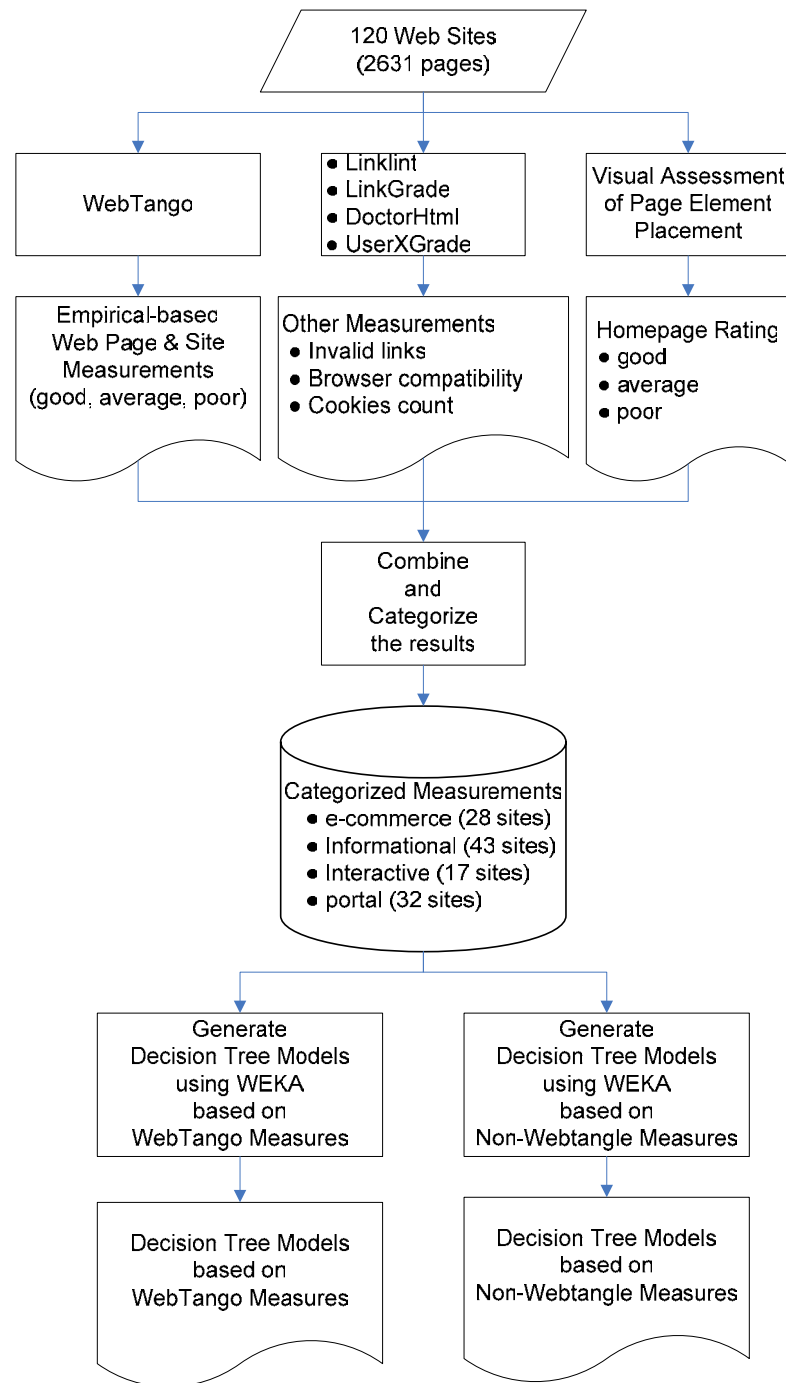


Figure 2.14. The overview of Clapsaddle's methodology (Clapsaddle, 2004).

1. IF ((Good Text Page Title Words <= 13) AND (Interactive Objects > 11) AND (SansSerif Font Count > 0) AND (Average Graphic Height > 10)) (30)
2. IF ((Script File Count <= 3) AND (HTML File Count <=2) AND (Average Graphic Height > 29) AND (Colored Body Words <= 232)) (107)
3. IF ((Script File Count <= 5) AND (Emphasized Body Word Count <= 550) AND (Download Time <= 48.7)) (46/1)

Figure 2.15. A decision tree model and rule set for “good” e-commerce Web sites (Clapsaddle, 2004, p. 88).

METRIC	ECOM		INFO		IACT		LINK	
	IG	RANK	IG	RANK	IG	RANK	IG	RANK
Script Size (bytes)	0.57419	1	0.46679	2	0.8114	2	0.7968	1
Page Width (pixels)	0.3547	2	0.64761	1	0.8789	1	0.1937	18
Bad Panel Color Combo Count	0.2047	3	0.1192	34	0.1676	56	0.2043	16
Maximum Font Size	0.20301	4	0.22223	8	0.5236	3	0.0678	62
Object Count	0.18464	5	0.24798	4	0.3185	21	0.116	41
Page Title Word Count (>64)	0.18003	6	0	0	0.0483	79	0.1017	48
MetaTag Word Count	0.17993	7	0.14881	21	0.1692	55	0.2185	12
Table Count	0.16983	8	0.23118	6	0.3648	16	0.2652	6
Text Title Score	0.159	9	0.05613	66	0.1704	54	0.0267	88
All Page Text Score	0.15855	10	0.21826	10	0.1555	59	0.1393	32
Colored Body Word Count	0.15225	11	0.11172	37	0.3269	20	0.2871	5
Average Graphic Height	0.14671	12	0.0341	82	0.4003	10	0.1041	47
SansSerif Font Count	0.14009	13	0.23157	5	0.2583	36	0.0481	72
Weblint Errors	0.13998	14	0.10268	42	0.3908	13	0.2378	9
Link Color Count	0.1351	15	0.20731	11	0.3958	12	0.131	38
SansSerif Word Count	0.13311	16	0.08996	45	0.2872	26	0.0473	74
Script File Count	0.12897	17	0.16541	18	0.3152	22	0.0636	64
Graphic File Count	0.12794	18	0.19142	12	0.2975	24	0.1692	23
Visible Page Text Score	0.1272	19	0.14164	24	0.1952	47	0.1993	17
Most Used Font Face	0.12574	20	0.03994	78	0.1954	46	0	0
Text Column Count	0.11794	21	0.13427	30	0.2284	42	0.2083	15
Good Page Title Word Count (>64)	0.11502	22	0.02202	87	0.0349	83	0.062	65
Redundant Graphic Count	0.11107	23	0.29907	3	0.2453	40	0.2921	4
Good Text Color Combo Count	0.10966	24	0.13858	27	0.1386	62	0.2185	12
Browser-Safe Color Count	0.10505	25	0.14176	23	0.1128	65	0.0753	59

Figure 2.16. The top measures (Clapsaddle, 2004, p. 83).

In summary, a significant amount of research has been conducted related to WebTango. The WebTango tool appears to be the most validated and most studied automated usability evaluation tool available today, and the documentation related to the

WebTango creation process is more widely available than for any other automated usability evaluation tool.

Web-based Learning

Web-based learning refers to any educational or learning environment that is delivered or connected via the Web. Many other terms have been used for an educational or learning environment that is partly or entirely Web-based, including the following: computer-based educational system (Granic, Glavinic, & Stankov, 2004), computer-supported collaborative learning (CSCL) (Kirschner, Strijbos, Kreijns, & Beers, 2004), distance learning, e-learning (Granic, Glavinic, & Stankov, 2004), interactive learning system (Reeves & Hedberg, 2003), online learning (Kearsley, 2005), virtual learning environment (VLE) (Chalk, 2002), Web-based educational system (Granic, Glavinic, & Stankov, 2004), Web-based instruction (WBI) (Khan, 1997), Web-based learning environment (Web-LE) (S.-K. Wang, 2003), Web-based training (WBT) (Khan, 2001), and Web-oriented educational system (Granic, Glavinic, & Stankov, 2004). For the purpose of consistent communication, *Web-based learning* has been used throughout this dissertation to refer to any of the above (and other) Web-based learning related labels.

Information-centric Web-based Learning

Web-based learning is a broad term to say the least. For better understanding, conceptual frameworks have been proposed to classify Web-based learning (Harmon & Jones, 1999; S.-K. Wang, 2003). The next two sections describe these two Web-based learning frameworks.

Harmon and Jones' Framework. Harmon and Jones (1999) developed a framework consisting of multiple levels of Web use in education: no Web use,

information Web use, supplemental Web use, essential Web use, communal Web use, and immersive Web use. This framework was intended to help educators to decide which type(s), if any, of Web usage they should employ in their teaching activities.

- *Level 0*, no Web use—as the name suggests, the Web is not used at this level.
- *Level 1*, information Web use—instructors post administrative material, a course syllabi, for example, on a Web site.
- *Level 2*, supplemental Web use—instructors provide added course material, such as magazine articles, on the Web to supplement core course content.
- *Level 3*, essential Web use—instructors post most, if not all, core course content on a Web site. In essence, students cannot be productive without regularly accessing the course Web site.
- *Level 4*, communal Web use—instructors and students interact with each others via Internet-based tools, such as e-mail, chat rooms, and bulletin boards. The communication tools are used to supplement face-to-face instructions.
- *Level 5*, immersive Web use—a course is conducted entirely online. Instructors and students communicate via the Internet. Content is posted, accessed, and exchanged via the Web.

Wang's Framework. Wang (2003) derived a framework for conceptualizing Web-based learning which consists of three levels: *Level 1*, utilizing communication tools supporting online community; *Level 2*, accessing instructional materials; *Level 3*, utilizing multimedia online applications for teaching and learning.

Level 1 refers to either commercially or freely available software programs that facilitate various forms of human communication, such as e-mail, instant messaging, Internet phones, video conferencing, bulletin boards, and blogs (see Figure 2.17). Minimal technical knowledge is required for setting up or using these communication tools, and all participants contribute content, instead of having an instructor or instructional designer be the primary content creator.

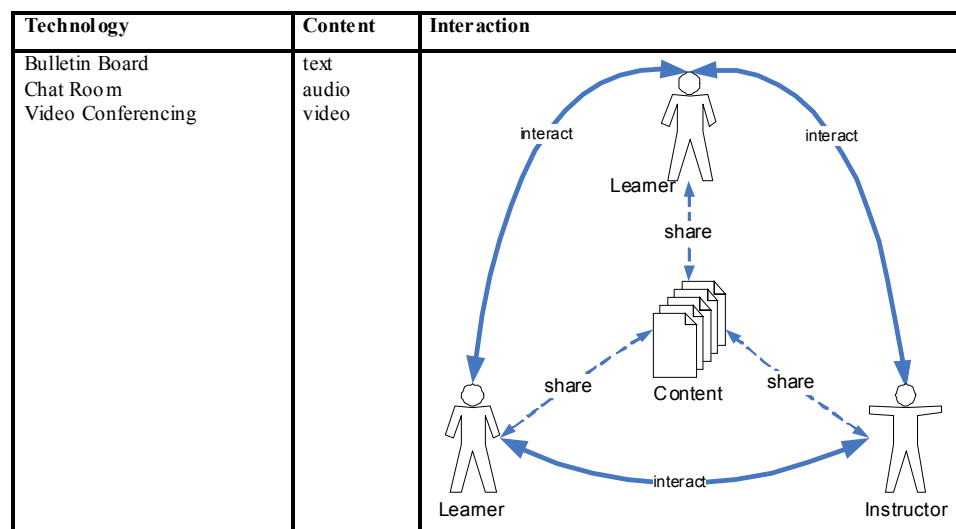


Figure 2.17. Wang's Level 1 Web-based learning.

Level 2 Web sites are either static (e.g., html) or dynamic (e.g., php or jsp pages) sites that are mainly for disseminating information (see Figure 2.18). A *level 2* Web site may be built by an instructor, instructional designer, or by a multidisciplinary team that consists of instructors, *subject matter experts* (SMEs), instructional designers, art designers, programmers, and evaluators. Users or learners usually access the site content.

Level 3 Web sites are sites that are embedded with highly interactive software programs or components (see Figure 2.19). *Level 3* Web-based learning sites mainly support interactive learning activities or tasks, as opposed to content dissemination alone.

Software design and programming knowledge are typically necessary for developing a *high interaction* Web-based learning site.

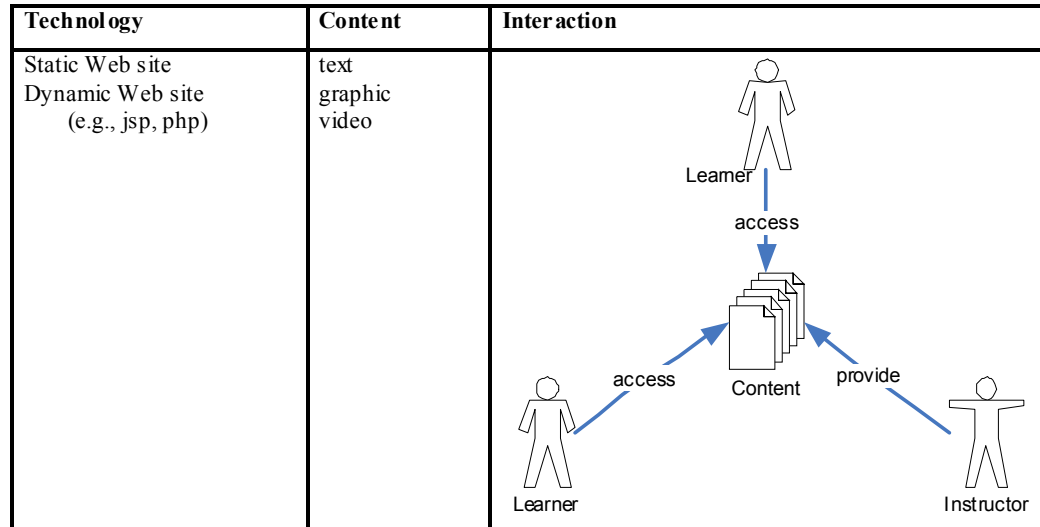


Figure 2.18. Wang's Level 2 Web-based learning.

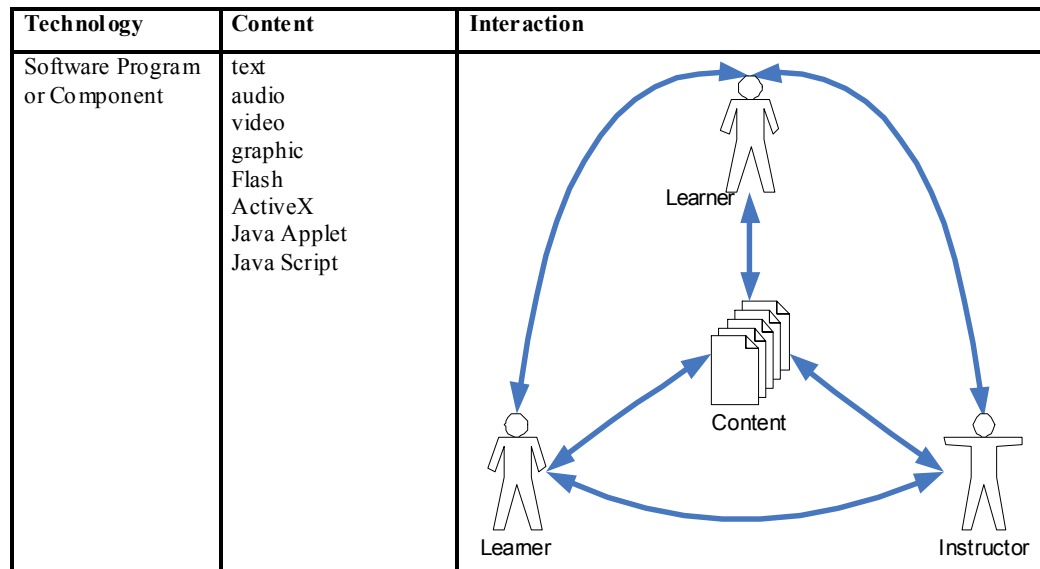


Figure 2.19. Wang's Level 3 Web-based learning.

Information-centric Web Interfaces. Ivory (2001) defined this type of Web interface as being *information-centric* when they were: “sites whose primary tasks entail

locating specific information” (p. 81). *Information-centric* Web interfaces mostly contain either static Web pages (e.g., html) or dynamic Web pages (e.g., php, jsp) (Ivory, 2001). Moreover, Ivory (2001) associated these *information-centric* interfaces with educational Web sites. The interfaces of several Web based learning types are information-centric: Web use (Harmon & Jones, 1999), supplemental Web use (Harmon & Jones, 1999), essential Web use (Harmon & Jones, 1999), as well as Wang’s (2003) “Level 2”. This dissertation focused on *information-centric* Web-based learning sites. Although I have been unable to find a reliable analysis, it seems plausible that the majority of Web sites deemed educational at this time are *information-centric*.

Web-based Learning Usability Evaluation

Web-based learning usability evaluation is a form of *formative evaluation* which may be viewed as a critical prerequisite to determining the overall effectiveness of learning in an instructional environment (Hughes & Burke, 2001; Reeves & Hedberg, 2003). Web-based learning evaluation as a comprehensive process involves not only usability factors, but also educational or pedagogical factors to achieve learning (Barab, MaKinster, Moore, & Cunningham, 2001; Clark & Mayer, 2003; Gagné, Wager, Golas, & Keller, 2005; Kirschner, Strijbos, Kreijns, & Beers, 2004; Mayer, 2001; S.-K. Wang, 2003). Such factors may include learning principles, instructional sequence, human cognition, social or group dynamics, motivation, and task ownership.

Sometimes a technological usability guideline contradicts a pedagogical principle. For instance, a widely accepted and empirically supported guideline says, “Provide a text equivalent for every non-text element that conveys information” (Koyani, Bailey, & Nall, 2004, p. 24). However, if the pedagogical goal is to enable learners to interpret

knowledge, converting from pictures to words, for example (L. W. Anderson, Krathwohl, & Bloom, 2001), this pedagogical goal contradicts with the usability guideline. When this type of dilemma occurs, a decision must be made as to whether the Web-based learning should follow the usability or pedagogical guidelines. In short, a Web-based learning designer must achieve a balance between conforming to both usability and pedagogical guidelines, and selecting only the most appropriate guideline.

While acknowledging the importance of pedagogical factors in Web-based learning, this dissertation focused on usability factors. Indeed, the usability quality of Web-based learning may be considered a prerequisite to pedagogical quality, as “un-usability” may prevent successful learning even if a Web-based learning is pedagogically sound (Kirschner, Strijbos, Kreijns, & Beers, 2004; Miller, 2005; Storey, Phillips, Maczewski, & Wang, 2002; S.-K. Wang, 2003).

Web-based Learning Usability Related Studies

Usability is considered important in Web-based learning, and accordingly some educational researchers have focused their attention on this challenge. Several Web-based learning usability related studies are highlighted in this subsection to provide a sense of the scope of such studies. Lohr *et al.* (2003) evaluated the usability of 11 Web sites containing self-paced instruction for two 1-credit technology courses. These courses consisted of self-paced instruction (Web-based), instructor-led workshops, and open labs with one-on-one assistance. This usability study followed the *analysis, synthesis, evaluation* and *change* (ASEC) model (Reigeluth & Nelson, 1997) for instructional design. The usability evaluation related question of the study was “Did the ASEC model lead to prototype improvement, as measured by a change in usability perceptions?” (Lohr

et al., 2003, p. 44). The three-semester long usability study analyzed data provided by 570 preservice teachers as learners and seven instructional designers. The instruments used for the study included journals of instructional designers and 43 five-point Likert scale question items for identifying the effectiveness, efficiency, and appeal of the Web sites. The results of the usability tests from each iteration (semester) were responsible for increasing usability scores of enhanced Web sites for subsequent semesters. Lohr *et al.* (2003) concluded that using a usability study as part of the ASEC model helped identify areas for improvement and areas where students were highly satisfied.

Wang *et al.* (2003) developed *Knowledge Innovation for Technology in Education* (KITE): a Web-based *case-based reasoning* (CRB) (Kolodner, 1993; Schank, 1982, 1998; Schank, Kass, & Riesbeck, 1994) knowledge repository to support a learning community. The researchers employed usability evaluation as part of an ongoing formation evaluation of the project. Wang et al. (2003) posted the following research question for the usability evaluation: “What is the usability level of the KITE knowledge repository search functions and navigation tools?” (p. 56). The instruments used for the usability study included observations, interviews, and a 10-point scale evaluation form that included nine measurement dimensions: 1) ease of use, 2) navigation, 3) level of confusion, 4) orientation within the system, 5) screen design, 6) information presentation, 7) media integration, 8) aesthetics, and 9) overall functionality (Reeves & Harmon, 1993). Fifty-two participants were involved in the usability evaluation. The usability testing results were used to improve the CRB engine of KITE.

Wang et al. (2003) stated that the usability portion of this study’s formative evaluation intended to determine whether KITE is easy to use, users are able to find the

information they want, and the search tools work. However, Wang's research did not provide more detailed information, such as specific questions that were addressed, usability evaluation methods that were used, or how the usability evaluation was implemented.

Granic et al. (2004) conducted a usability evaluation on a previously developed Web-based intelligent authoring shell or a Web-based educational system. The objectives of the study were to 1) identify the usability shortcomings, 2) identify the expected improvements, and 3) determine whether the evaluation approach used in the study is a promising one for other Web-based education systems (Granic, Glavinic, & Stankov, 2004). The instruments used in the study were a scenario-based usability test, a guideline evaluation, and a usability questionnaire. Five developers of the Web-based educational system and 10 computer science, mathematics, and polytechnic undergraduate students were the participants of the usability study. The results indicated that "useful usability validation with significant identification of inherent interface weaknesses can be performed quite easily and quickly, with relatively no cost except the employees' time" (Granic, Glavinic, & Stankov, 2004, Conclusion section, ¶ 3). These results have been used to improve the newer versions of Web-based educational systems.

Granic's study appears to have been very properly conducted. The objective was clearly articulated. Empirical and heuristic methods were used to measure observable issues such as task performance and error rates. A questionnaire was used to uncover unobservable aspects of a user interface, such as users' satisfaction or frustration. The methodology and results were sufficiently described.

Storey *et al.* (2002) conducted a study comparing two commercially available Web-based learning tools: WebCT 2.0 (WebCT, 2005) and Blackboard 6.0 (Blackboard, 2005). (In 2005, newer versions of these tools than the versions used in the study were released.) The study intended to find out 1) the usability rating from students, 2) the required effort from students to learn to use the tools, 3) the usability rating from the instructors and administrators, 4) the impact of learning, and 5) the students' feelings about deploying these tools in university courses. Fifty-four computer science students, one instructor, one system administrator and three research assistants participated in this one-semester long study by filling out online questionnaires. Storey *et al.* (2002) reviewed the usability principles (Norman & Draper, 1986) and found that “most of these usability principles were violated by the tools we evaluated and negatively impacted students and their attitudes towards these tools” (p. 92). This study resulted in many recommendations for universities to improve usability of Web-based learning tools based on feedback from students, instructors, and system administrators. Storey *et al.* (2002) stated that questionnaires were used to gather information concerning these areas: the ease of learning, ease of use, students' learning curve for the tool, and usefulness of the tool. The researchers did not describe any survey research methodology as the basis for designing the questionnaires, nor did they reveal whose “areas of concerns” seem to be at a very high-level. Based on information given by the researchers, it is difficult to determine the validity and reliability of the questionnaires.

Sheard and Markham (2005) evaluated a Web-based learning site's interfaces by following the *trailing* research framework. *Trailing* research (Finne, Levin, & Nilssen, 1995) combines formative and summative evaluation to aim for high *credibility* and high

relevancy of research (Sheard & Markham, 2005). The trailing research framework 1) encourages high credibility by suggesting that external evaluators work cooperatively with stakeholders, and 2) encourages high *relevancy* by suggesting “the evaluation approach and outcomes reflect the interests and needs of the stakeholder” (Sheard & Markham, 2005, p. 355).

Sheard and Markham (2005) conducted two rounds of studies. The study participants almost entirely consisted of undergraduate students enrolled in Industrial Experience Project courses of a Bachelor of Computing unit (Sheard & Markham, 2005). The *first round* of the study employed three online surveys, one paper-based survey, and a log analysis involving 157 study participants. The online surveys, consisting of a five-point Likert scale and open-ended question items, intended to uncover several aspects of Web interfaces: navigation, reliability, response time, available resources, and the usefulness of each resource. The log file analysis kept track of participants’ navigation paths. However, the “huge volume of data” (Sheard & Markham, 2005, p. 361) generated by the analysis was unsuitable for leading to meaningful findings. Sheard and Markham (2005) acknowledged these disappointments on this round of study: the response rate of online surveys and open-ended question items was poor; students were dissatisfied with the Web site, collected data did not provide information suitable for addressing the usability issues, and the Web site being studied was complex, perhaps contributing to the disappointing results. The researchers made no mention of creating surveys based on a survey research methodology; thus it is not too surprising their surveys generated low-quality results. In essence, their research basically showed what not to do in a Web usability evaluation study.

The second round of the Sheard and Markham (2005) study employed observations, interviews, and two paper-based surveys (pre and post observations) involving 18 study participants. Each participant was observed while he or she performed tasks. The researchers did not state whether the participants had pre-determined tasks or could access the site freely. The participants filled out one paper-based survey prior to their observation session; afterward, they were interviewed and filled out a paper-based post-observation survey. The results of this second round of study were meaningful enough to serve as the basis of a major reconstruction of the Web site being studied. The researchers attributed this success to the “intensive[ness]” (Sheard & Markham, 2005, p. 363) of this round of the study. The surveys, observations, and interviews occurred in a captive environment, as opposed to filling out online surveys at participants’ convenience. The second round of the study appears to be much more focused than the first round. Although the researchers did not describe the research design and procedure in detail, they seem to have been much more mindful of research quality than they were at the beginning of the study. The researchers did not explicitly discuss research design, validity, and reliability. Thus, the quality of this research could have been even higher, if the principles of good research design were consciously followed, and the validity and reliability of the research instruments were clearly addressed.

Evans and Sabry (2003) developed a heuristic evaluation model, a three-way model of interactivity (3-WMI), and conducted a study using 3-WMI to evaluate three Web-based learning sites. The 3-WMI model organizes computer-learning interactions into three types or “ways”: *initiation*, *response*, and *feedback*. The three-way model

provides a framework for heuristic evaluation of Web-based learning sites. The 3-WMI model has the dual-goal of guiding or training evaluators to follow the framework, and to generate new heuristics. A study of the 3-WMI model started with developing a provisional set of nine heuristics that are “comprehensive but not necessarily exhaustive” (Evans & Sabry, 2003, p. 91): making navigation easy, engaging learner frequently, allowing for reflection, using a variety of interactions, applying what has been taught, initiating interactions, allowing for learner response, and providing feedback. After these heuristics were developed, six expert evaluator participants independently evaluated three Web-based learning sites following the 3-WMI model and based on the set of nine heuristics. Finally, the results from all participants were combined and analyzed. The researchers concluded that the results of the study “led to reconsideration of the set of nine heuristics” (Evans & Sabry, 2003, p. 97); moreover, they stated that the process of using 3-WMI to evaluate Web-based learning “is easily generalized to allow the evaluation of any form of interactive multimedia system” (Evans & Sabry, 2003, p. 98)

At first glance, the 3-WMI model seems to be an interesting and perhaps promising framework for heuristic evaluation. The quantitative data analysis was rigorously conducted. However, the basis of the study seems to be weak. *First*, the provisional set of nine heuristics was problematic. There was little discussion about the methodology for developing the provisional set of nine heuristics. There was no citation associated with any one of the nine heuristics. Most of the heuristics display the typical symptoms of poor Web usability heuristics: they were vague to the point of being meaningless and were difficult to apply meaningfully (Ivory, 2001). *Second*, a pre-evaluation workshop was conducted to “motivate” (Evans & Sabry, 2003, p. 92) the

participants in using 3-WMI. It is reasonable to believe that this motivational workshop may have introduced biases into the study. *Finally*, the researchers' conclusions seem to be overly simplistic and optimistic.

In summary, usability evaluation is considered important in Web-based learning. Research studies related to manual usability evaluation and Web-based learning are fairly typical. There have been several evaluations of educational Web sites using automated usability or accessibility tools reported online (Kelly, 2002a, 2002b, 2003a, 2003b). However, these evaluations do not constitute formal research studies. Research studies on automated usability evaluation tools and Web-based learning are lacking.

Conclusion

Although Web sites share many of the characteristics of traditional computer software, Web site usability has so many distinctions from *software usability* that Web site usability is considered a sub-area within the term *usability*. Web-based learning usability not only inherits Web usability factors, but also includes pedagogical factors. Despite the beliefs of Web site builders' in the value of usability evaluation, both general Web sites and Web-based learning sites usually have low usability quality. Conducting a formal usability test can be time and resource intensive. Alternative usability evaluation methods such as heuristic evaluation and automated usability evaluation have been used to either positively supplement or negatively substitute for formal usability testing to reduce time and cost. Automated usability evaluation has shown promise in helping Web site builders to improve usability quality, especially for site builders who don't have formal training in usability design and evaluation. However, there appears to be a lack of

research studies related to the use of automated usability tools in designing and evaluating the user interfaces of Web-based learning resources.

The next chapter describes the two parts of this study's methodology: 1) a usability evaluation of Web-based learning sites using WebTango, and 2) the survey study of the perception of Web-based learning site builders.

Chapter 3: Methodology

As described in Chapter 1, the two major goals of this dissertation are as follows:

to gain insight into the current state of Web-based learning sites' usability based on an automated evaluation tool, and to understand the perceptions of Web-based learning site builders on the usefulness of usability analysis results generated by an automated tool.

To fulfill these goals, the methodology of this dissertation had two phases: 1) a usability analysis of Web-based learning sites using WebTango, and 2) a survey study of Web-based learning site builders (see Figure 3.1). A pilot study was conducted prior to the survey study being implemented.

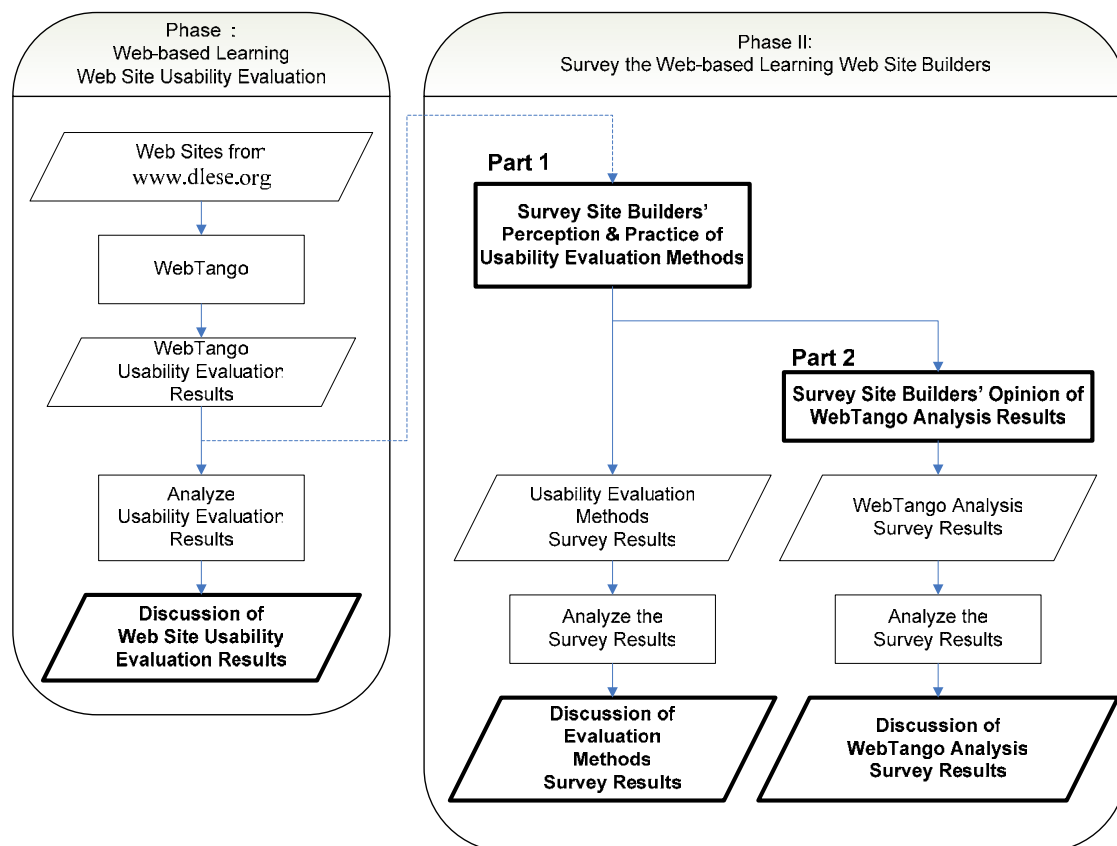


Figure 3.1. An overview of methodology.

This chapter describes the research methodology I designed and implemented to conduct a usability evaluation using WebTango as well as the survey study. WebTango was selected for several reasons. *First*, WebTango is an automated evaluation tool that is based on empirical measures. *Second*, WebTango offers a more extensive set of measures specific to usability than any other *guideline review* usability evaluation tool available today. *Third*, heuristic evaluation (or *guideline review*) is an accepted method to supplement formal usability testing. *Fourth*, WebTango is a *guideline review* usability evaluation tool intended for *information-centric* Web sites and so is suitable for many educational Web sites (Ivory, 2001). *Finally*, automated usability evaluation tools are particularly suited for Web site builders who are not formally trained in usability design and testing.

Phase I: Web-based Learning Site Usability Evaluation

As described in Chapter 1, the first research question of this study is this:

1. What is the usability quality of Web-based learning sites according to an automated usability evaluation tool?

This research question was addressed by using WebTango to analyze Web-based learning sites. The analysis included the following items related to Web usability supported by WebTango: *text element*, *text formatting*, *graphic element*, *graphic formatting*, *link element*, *link formatting*, *page title*, and *page formatting* (see Table 3-1 and Appendix B: The WebTango Web Interfaces Usability Measurements). Each site was analyzed individually.

Table 3-1

Usability Evaluation Items Supported by WebTango and Included in this Study

Evaluation Item	Short Description
Text Element	“Amount of text, type, quality, and complexity. Includes visible and invisible text” (Ivory, 2003, p. 71).
Text Formatting	“How body text is emphasized; whether some underlined text is not in text links; how text areas are highlighted; font styles and size; number of text colors; number of times text is repositioned” .
Graphic Element	“Number and type of images” (Ivory, 2003, p. 71).
Graphic Formatting	“Minimum, maximum, and average image width and height; page area covered by images” (Ivory, 2003, p. 71).
Link Element	“Number and type of links” (Ivory, 2003, p. 71).
Link formatting	“Colors used for links and whether there are text links that are not underlined or colored” (Ivory, 2003, p. 71).
Page Title	The use of "descriptive and different title on each page" (Koyani, Bailey, & Nall, 2004, p. 76) on a Web site.
Page Formatting	“Color use, fonts, page size, used of interactive elements, page style control, and so on” (Ivory, 2003, p. 71).

The Target Web Sites

A usability evaluation, using WebTango, was performed on Web sites from the *DLESE Reviewed Collection (DRC)* of the *educational resources* section (DLESE, 2005d) in the Digital Library for Earth System Education (DLESE) (<http://dlese.org/>) (DLESE, 2005a). DLESE is a component of the National Science Digital Library (<http://nsdl.org/>). DLESE is described as follows:

a distributed community effort involving educators, students, and scientists working together to improve the quality, quantity, and efficiency of teaching and learning about the Earth system at all levels.... DLESE resources include electronic materials for both teachers and learners, such as lesson plans, maps, images, data sets, visualizations, assessment activities, curriculum, online courses, and much more (DLESE, 2005a, Overview of DLESE, ¶ 1).

A screen capture of the main screen for DLESE is shown in Figure 3.2.

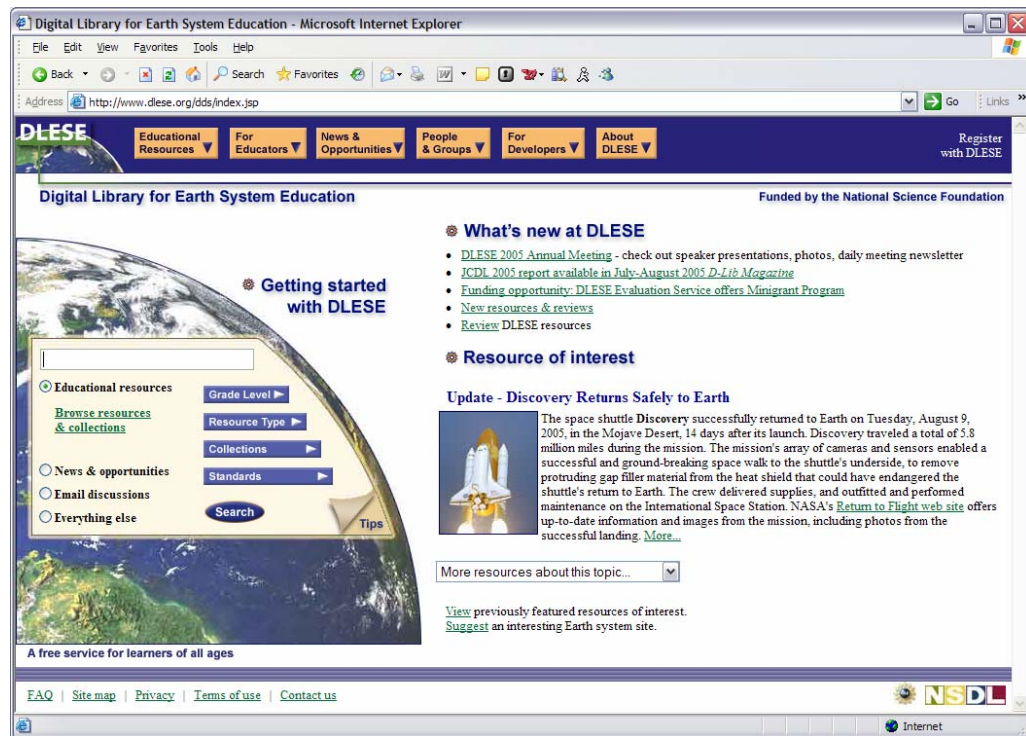


Figure 3.2. A screen sample of DLESE (DLESE, 2005b).

The *educational resources* section of DLESE links to more than 10,000 Web-based learning resources (Web sites) that include “lesson plans, scientific data, visualizations, interactive computer models, and virtual field trips—in short, any Web-accessible teaching or learning material” (DLESE, 2005c, ¶ 1). DLESE is noteworthy not only because it is recognized as a groundbreaking digital library (Wright, Marlino, & Sumner, 2002), but because it addresses earth system science education, a focus that is increasingly important for global scientific literacy (Barstow & Geary, 2002).

The DRC is a special subset of DLESE that links to nearly 600 Web sites that meet seven review criteria: “high scientific accuracy, good pedagogical effectiveness, ease of use, clarity and completeness of documentation, motivating for learners, show robustness, and illustrate significance of content” (DLESE, 2005d). The DRC sites were selected as the focus for this study for two major reasons: 1) This site collection consists

of a reasonably large number of sites for analysis (in comparison to other WebTango related studies), and 2) The sites have been evaluated by DLESE reviewers and have met seven criteria including pedagogical effectiveness and ease of use.

Instead of the entire 600 DLESE-DRC sites, only 130 Web sites in the DRC could be analyzed for various reasons. The following list highlights the major reasons of excluding Web sites from this study:

- Web sites that can not be analyzed by WebTango. For example, pdf file, graphic file, Macromedia Flash component, JavaScript component, etc.
- Web sites or pages that merely describe an actual learning site or a Web site solely intended for ordering other types of learning material
- Learning material that is embedded in a large portal because the content is likely to be submitted via a standardized template (see Figure D.2 in Appendix D).
- Web sites that contain articles or reports rather than learning materials
- Web sites that were not able to be analyzed prior to September 19, 2005 when WebTango was suddenly and without prior notice deactivated.

Collection of Data and Analysis

The following list illustrates the high-level steps that were taken for analyzing Web-based learning sites found in the DRC. *First*, the Web addresses of resources in DRC were collected and saved into a file. *Second*, the Web addresses were submitted to the WebTango Analysis tool. *Third*, WebTango analysis result data files (one file per Web address) were downloaded. *Fourth*, descriptive statistics of WebTango results were

generated using SPSS (SPSS, 2001). *Finally*, the analysis results were reported and discussed.

Using WebTango, I personally analyzed 35 of the 130 DLESE-DRC sites. With the advice and support of Professor Thomas C. Reeves, Chun-Min Wang, a fellow UGA doctoral student who was also a graduate research assistant working with Professor Reeves on the overall evaluation of DLESE, analyzed the remaining 95 sites using WebTango. Figure 3.3 summarizes the methodology of *phase I* of this dissertation.

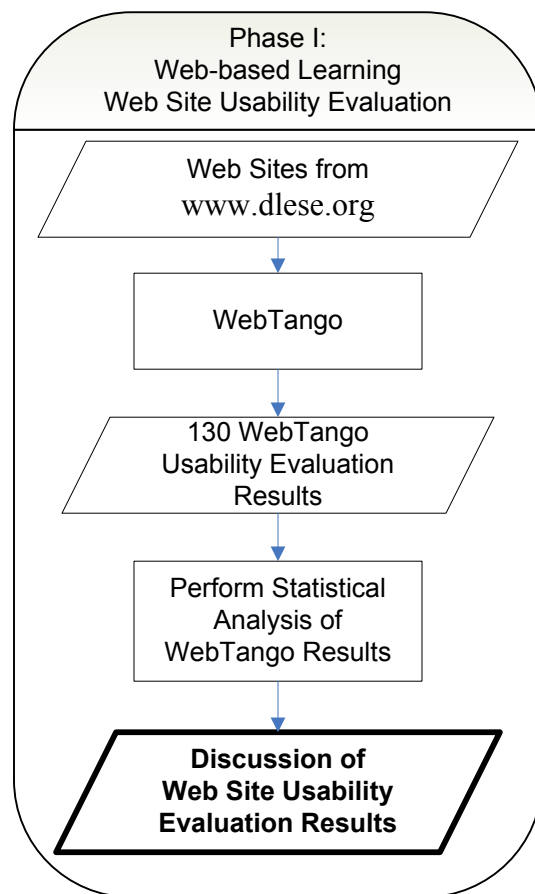


Figure 3.3. The methodology summary of phase I.

Phase II: Survey the Web-based Learning Web Site Developers

As described in Chapter 1, the other research questions of this study are as follows:

2. According to the perception of Web-based learning site builders,
 - 2.1. What is the degree of expertise among Web-based learning site builders regarding usability evaluation methods?
 - 2.2. How frequently do the Web-based learning site builders use usability evaluation methods?
 - 2.3. What is the *perception* among Web-based learning site builders regarding usability evaluation?
3. According to the perception of Web-based learning site builders, how useful are the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality?

Survey methods are the logical choice for addressing these type of questions (Isaac & Michael, 1995). The survey study consisted of two parts: *Part one* was designed to address research questions 2.1, 2.2 and 2.3. *Part two* was designed to address research question 3. Figure 3.4 provides an overview of the survey study.

Web-based Research Survey

This study utilized *cross-sectional surveys* (Babbie, 1973) implemented via the Web to *describe* and *explore*, at one point in time, the Web site builders' perceptions of WebTango analysis results of their own Web sites. A *cross-sectional* survey is appropriate for this study because, according to Babbie (1973), survey research may provide *descriptions* for "making descriptive assertions about some population" (p. 57),

and *exploration* for discovering findings beyond a research's preconceptions about a topic. Moreover, *cross-sectional* surveys involve data collection “at one point in time from a sample selected to describe some larger population *at that time*” (Babbie, 1973, p. 62).

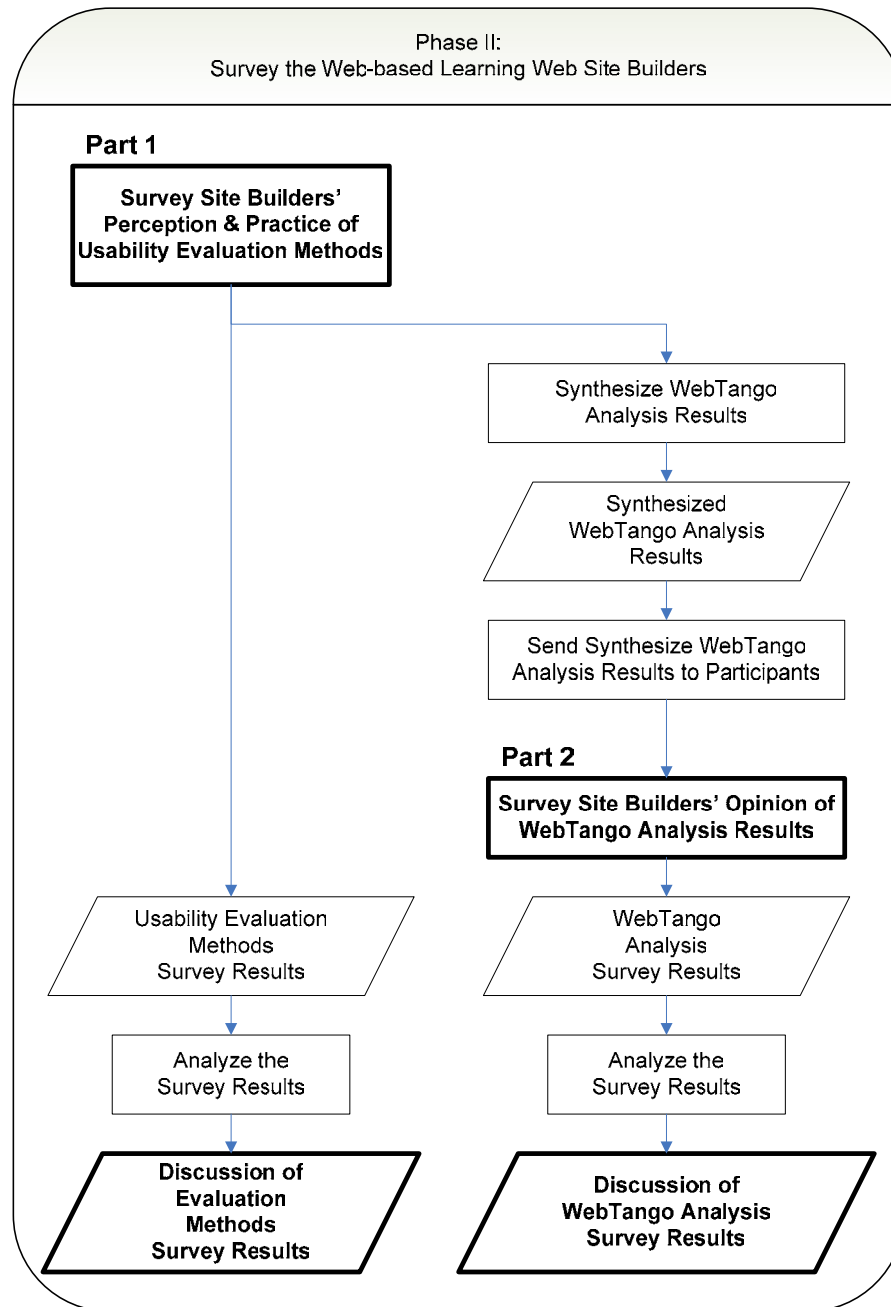


Figure 3.4. An overview of the survey study.

Administering this *cross-sectional* survey via the Web is also appropriate. According to T. Anderson and Kanuka (2003), a Web-based survey has the following potential advantages: cost savings, time savings, higher rates of return, increased accuracy, direct participant entry of data, design flexibility, and faster creating and delivery.

In summary, because of their many advantages, Web-based *cross sectional* surveys was the specific survey method to be used. The survey was divided into two parts: *Part one* addressed research questions 2.1., 2.2 and 2.3; the Web site builders' perception and practice of usability methods. *Part two* of the survey addressed research question 3; Web site builders' opinion of usability evaluation results generated by an automated tool.

Sample and Instrumentation

The survey sample to which the inquiry was addressed consisted of 35 people who identified themselves as the contact person (i.e., designer, owner, etc.) of 35 of the 130 Web sites included in *phase I* of this study (see Phase I: Web-based Learning Site Usability Evaluation section).

The survey instruments, consisting of *data collection questions* (Punch, 2003), were constructed by the researcher. The data collection questions design of these survey instruments were guided by 1) the research questions presented in Chapter 1 and earlier in this chapter, and 2) a synthesis of the research literature in Web usability and Web-based learning usability presented in Chapter 2. Survey instruments, the informed consent form, and synthesized WebTango analysis results are illustrated in Appendix E: Survey Instruments.

Collection of Data and Field Procedures

The following list illustrates the high-level steps that were taken in conducting the Web-based survey:

- Locate contact person for the sites via the telephone and e-mail.
- Create survey Web sites.
- Invite site contact person to participate in the *part one* survey study via e-mail.
- Send reminder e-mails to site contact persons who did not respond after three days.
- Stop accepting survey data after one week.
- Send a) the URL of synthesized WebTango analysis and b) the URL of part two survey to site contact person via e-mail.
- Send reminder e-mails to site contact persons who did not respond after three days.
- Stop accepting survey data after one week.
- Retrieve and analyze the data for both parts of the survey study.

As described below, identifying and recruiting more than 35 contact people proved to be impossible despite diligent efforts.

Nearly 300 contact persons of DLESE-DRC Web sites were contacted, via e-mail or telephone, to identify the appropriate participants for this study. The sites selected were potentially analyzable by WebTango. Fifty-five representatives of the 300 DLESE-DRC Web sites were identified as potential participants of the survey study, and 55 Web sites were included in the survey study. However, only 35 of the 55 Web sites were analyzed before WebTango was permanently deactivated in September, 2005. After

being sent survey invitations, followed by numerous e-mail and phone reminders, 30 people participated in the *Part One* survey, 29 participated in the *Part Two* survey.

Data Analysis

The results of the Likert scale format responses were reported as *verbal description, tabular form, and/or chart form* with descriptive statistics test results. The University of Georgia's Survey Research Center (2005), and the Academic Computing Center (2005), as well as various survey research guides (T. Anderson & Kanuka, 2003; Babbie, 1973; Dillman, 2000; Fowler, 2002; Schonlau, Fricker, & Elliott, 2002), were consulted to maximize the integrity of the data analysis. The Likert scale format responses were analyzed by using SPSS (SPSS, 2001). The open-ended responses were analyzed via textual content analysis.

Limitations

The usability evaluation inherits the general limitations of automated usability evaluation tools, and specific limitations of WebTango, as described in Chapter 2. Additionally, the study included Web sites from only one section of DLESE, i.e., the DRC. Whether the usability results for this subset of the DRC generalize to the entire DLESE collection of more than 10,000 resources is a question beyond the scope of this study. Unfortunately, WebTango server has been permanently deactivated since September 19, 2005, excluding additional DLESE-DRC sites from this study.

The survey phase of the study has the following limitations: The participants of this study were limited to builders of learning resources included in the DRC who could be identified and recruited to participate in the study. Web-learning resources can become "orphaned" for many reasons including the original designers moving onto other

jobs, retirement, name changes, and changes in contact information not being updated within the resource.

This survey also inherited some limitations of a Web-based Survey:

- Information overload (T. Anderson & Kanuka, 2003)—faculty members may ignore e-mail invitations for surveys because they are likely to receive a large number of e-mails.
- Procrastination (T. Anderson & Kanuka, 2003)—A Web-based survey can be responded to at anytime. This convenience may encourage procrastinators to postpone and result in their not responding.
- Challenges of enlisting cooperation (Fowler, 2002).

Finally, I designed the survey instruments and specified the data collection questions. The survey instruments were validated as recommended by survey design guidelines. To minimize the risk of using unproven survey instruments, a pilot study was conducted before the final survey study was implemented.

Summary

This study was conducted in two phases. *Phase I*, 130 of the Web-based learning sites linked to DLESE- DRC, a digital library, were evaluated using WebTango. These sites have been evaluated by DLESE reviewers and have met seven evaluation criteria including pedagogical effectiveness and ease of use. *Phase II*, a survey study was conducted to gain an understanding of the perceptions of Web-site builders on the usefulness of the WebTango analysis results in identifying the usability issues of their own Web sites. The survey also was designed to discern the site builders' perceptions,

and practices of Web usability. As described in the next chapter, a pilot study was conducted prior to implementing the full survey study.

Chapter 4: The Pilot Survey

A pilot study was conducted to identify issues related to the survey instruments and data analysis procedure. Twenty DLESE-DRC Web site builders were invited via e-mail to participate in this study, as opposed to the actual study, because their Web sites were not able to be analyzed using WebTango because WebTango was unexpectedly deactivated. The data collected during the pilot study were excluded from the actual study.

Prior to the pilot study, students enrolled in the Fall 2005 Instructional Product Evaluation (EDIT 8350) course taught by Professor Thomas C. Reeves at the University of Georgia performed a test-run of the survey Web site to identify mechanical issues. No problems were reported.

Two survey instruments were used in the pilot study. The first instrument was designed to address research questions #2.1, #2.2 and #2.3. The second survey instrument was designed to address research question #3. For the pilot study, 13 participants (65%) responded to the first survey instrument, and 10 participants (50%) responded to the second. As described below, the pilot study proved to be an invaluable activity that enhanced the actual study.

The Heuristic Evaluation

A heuristic evaluation was conducted on the 20 Web sites because WebTango, a Web-based tool, was unexpectedly permanently deactivated during this dissertation study. The heuristic evaluation can be considered an approximation of the WebTango analysis process. Eight WebTango measures were used for the evaluation: *text element*, *text formatting*, *graphic element*, *graphic formatting*, *link element*, *link formatting*, *page*

title, and *page formatting* (see Table 3-1 and Appendix B: The WebTango Web Interfaces Usability Measurements). Each measure was rated qualitatively and numerically as follows: *above average* (3), *average* (2), and *below average* (1). Text, graphic link, and page-related, research-based Web design and usability guidelines (Koyani, Bailey, & Nall, 2004) were used as basis for the heuristic evaluation. The intention of the heuristic evaluation was to generate the most realistic and consistent usability quality ratings for the pilot study, not to match or substitute WebTango ratings. The subsequent sections describe the results of the pilot study.

Results of Research Questions #2.1 and #2.2

As described in Chapter 1, and previously in this chapter, research questions 2.1 and 2.2 are as follows:

2. According to the perception of Web-based learning site builders,
 - 2.1. What is the degree of *expertise* among Web-based learning site builders regarding usability evaluation methods?
 - 2.2. How frequently do the Web-based learning site builders *use* usability evaluation methods?

This subsection describes the analysis results of participant responses to questions #1 through #6 of the first survey instrument that were intended to address research questions #2.1 and #2.2, as shown in Figure 4.1.

Descriptive statistics and *histograms* were used to describe the *central tendency*, *variability*, *size*, and *distribution* of the responses to survey questions #1 through #6 of the first instrument (see Figure 4.1).

Please check your **level of expertise** with the following [usability examination methods](#) using a scale where 1 = "Not familiar with it" to 5 = "expert in its application."

#1. Usability Testing: User(s) perform predetermined tasks while being observed and recorded.

#2. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.

#3. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.

Please check your **use** of the following [usability examination methods](#) using a scale where 1 = "Never" to 5 = "always."

#4. Usability Testing: User(s) perform predetermined tasks while being observed and recorded.

#5. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.

#6. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.

Figure 4.1. Survey questions #1 through #6 that addressed research questions #2.1 and #2.2.

The Descriptive Statistics

Table 4-1 illustrates the descriptive statistics of the scores of survey questions #1 through #6 of the first survey instrument in SPSS default output format. Table 4-2 illustrates the association between each survey question and its descriptive statistics.

On a scale of 1 (not familiar with it) to 5 (expert in its application), the level of expertise in usability methods of the survey participants was 2 (not very familiar with it). On a scale of 1 (never) to 5 (always), the participants almost never use usability methods. These results indicated the following:

- The degree of expertise among Web-based learning site builders regarding usability evaluation was relatively low (2 on a scale of 5 where 1 is lowest and 5 is the highest).
- Web-based learning site builders used usability evaluation methods infrequently (2 on the scale of 5 where 1 is lowest and 5 is the highest).

Table 4-1

The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #1 through #6 of the First Survey Instrument in the Pilot Study

Statistics		Q1	Q2	Q3	Q4	Q5	Q6
N	Valid	13	13	13	13	13	13
	Missing	8	8	8	8	8	8
Mean		2.54	2.38	1.69	2.00	2.08	1.46
Median		3.00	3.00	2.00	2.00	2.00	1.00
Mode		3	3	1	1	1	1
Std. Deviation		.967	1.121	.751	1.155	1.382	.967
Skewness		-.780	-.079	.611	.768	1.183	2.085
Std. Error of Skewness		.616	.616	.616	.616	.616	.616
Minimum		1	1	1	1	1	1
Maximum		4	4	3	4	5	4

Figure 4.2 illustrates the mean scores of survey questions #1 through #6 of the first survey instrument.

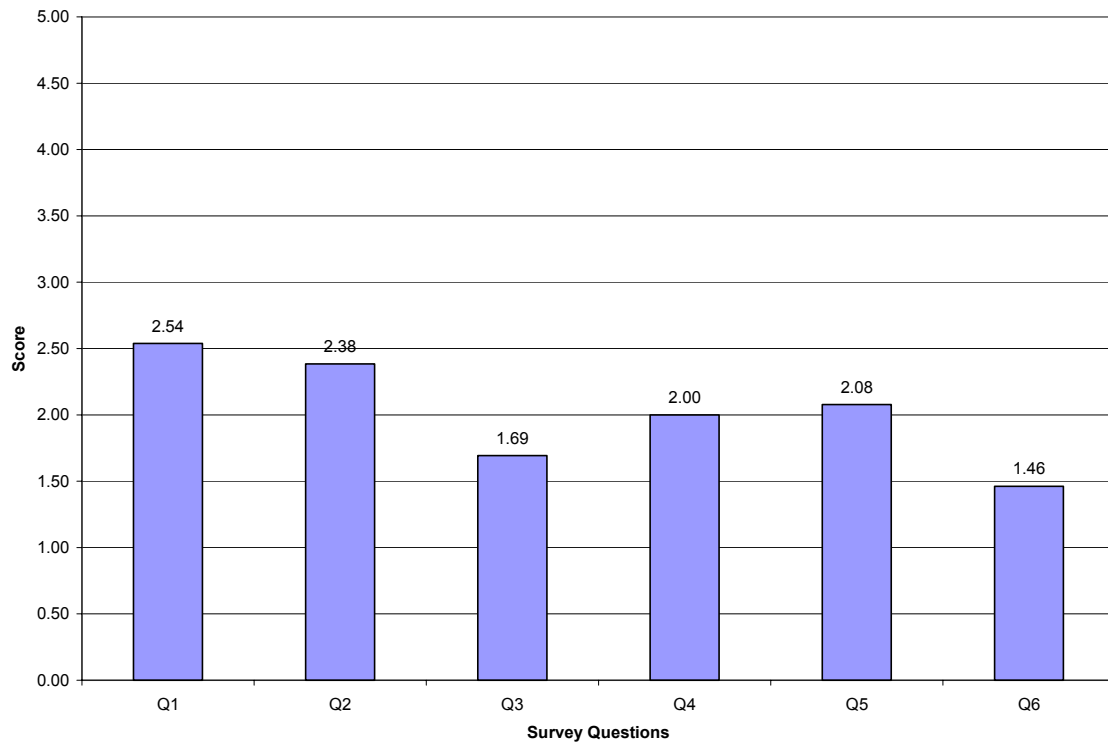


Figure 4.2. The mean scores of survey questions #1 through #6 of the first survey instrument in the pilot study.

Table 4-2

The Descriptive Statistics of Survey Questions #1 through #6 of the First Survey Instrument in the Pilot Study

Survey Questions	N	Missing	Mean	Median	Mode	SD	Skew	Min	Max
#1. Usability Testing: User(s) perform pre-determined tasks while being observed and recorded.	13	8	2.54	3.00	3	.967	-.780	1	4
#2. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.	13	8	2.38	3.00	3	1.121	-.079	1	4
#3. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.	13	8	1.69	2.00	1	.751	.611	1	3
#4. Usability Testing: User(s) perform pre-determined tasks while being observed and recorded.	13	8	2.00	2.00	1	1.115	.768	1	4
#5. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.	13	8	2.08	2.00	1	1.382	1.183	1	5
#6. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.	13	8	1.46	1.00	1	.967	2.085	1	4

Results of Research Question #2.3

As described in Chapter 1, the other research questions of this study are as follows:

2. According to the perception of Web-based learning site builders,

2.3 What is the *perception* among Web-based learning site builders regarding usability evaluation?

This subsection describes the analysis results of participant responses to questions #7 through #12 of the first survey instrument, which address research question #2.3, as shown in Figure 4.3.

Please check your agreement with each of the following statements using the scale below.
 1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

#7. I am interested in further improvement of the usability of my Web site.

#8. Usability is an essential feature of online learning resources.

#9. Usability is usually overlooked in the design of online learning resources.

#10. Personally, I would like to know more about usability in the context of online learning resources.

#11. It would be very helpful to automate the process of finding usability problems in online learning resources.

#12. Personally, I am skeptical about the value of automating the process of finding usability problems in online learning resources.

Figure 4.3. Questions #7 through #12 of the first survey instrument which addressed research question #2.3.

Descriptive statistics and histograms were used to describe the *central tendency, variability, size, and distribution* of the responses to survey questions #7 through #12 of the first instrument (see Figure 4.3).

The Descriptive Statistics

Table 4-3 illustrates the descriptive statistics of the scores of survey questions #7 through #12 of the first survey instrument in SPSS default output format. Table 4-4 illustrates the association between each survey question and its descriptive statistics.

Table 4-3

The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #7 through #12 of the First Survey Instrument in the Pilot Study in the pilot study

Statistics		Q7	Q8	Q9	Q10	Q11	Q12
N	Valid	13	13	13	13	13	13
	Missing	8	8	8	8	8	8
Mean		4.15	4.85	3.31	4.38	3.77	2.92
Median		5.00	5.00	3.00	4.00	4.00	3.00
Mode		5	5	3	4 ^a	4	2 ^a
Std. Deviation		1.214	.376	.630	.650	.832	.954
Skewness		-1.662	-2.179	-.307	-.572	-.528	.854
Std. Error of Skewness		.616	.616	.616	.616	.616	.616
Minimum		1	4	2	3	2	2
Maximum		5	5	4	5	5	5

a. Multiple modes exist. The smallest value is shown

Figure 4.4 illustrates the mean scores of survey questions #7 through #12 of the first survey instrument in the pilot study.

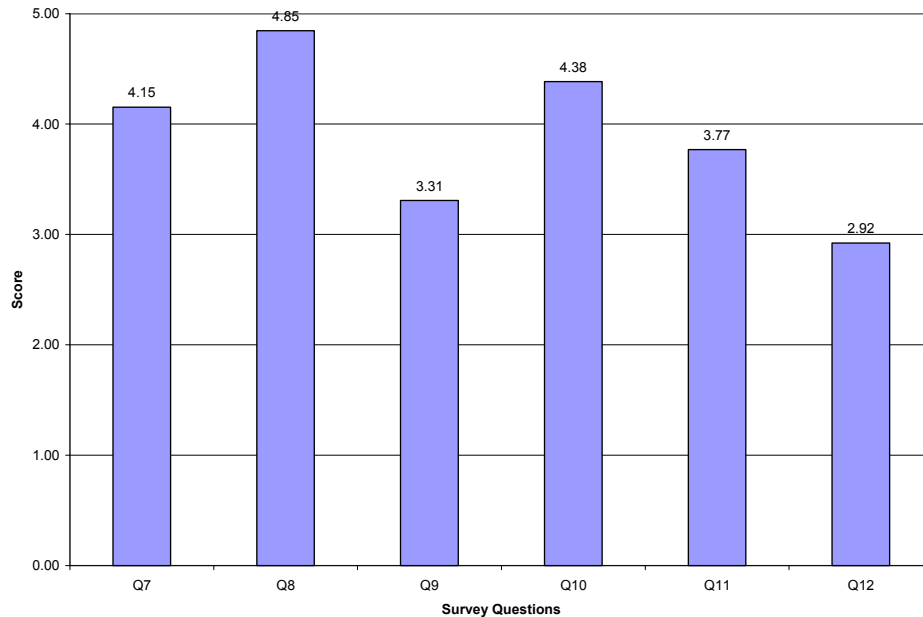


Figure 4.4. The average scores of survey questions #7through #12 of the first survey instrument in the pilot study.

Table 4-4

The Descriptive Statistics of Survey Questions #7 through #12 of the First Survey Instrument in the Pilot Study

Survey Questions	N	Missing	Mean	Median	Mode	SD	Skew	Min	Max
#7. I am interested in further improvement of the usability of my Web site.	13	8	4.15	5.00	5	1.214	-1.662	1	5
#8. Usability is an essential feature of online learning resources.	13	8	4.85	5.00	5	.376	-2.179	4	5
#9. Usability is usually overlooked in the design of online learning resources.	13	8	3.31	3.00	3	.630	-.0307	2	4
#10. Personally, I would like to know more about usability in the context of online learning resources.	13	8	4.38	4.00	4	.650	-.572	3	5
#11. It would be very helpful to automate the process of finding usability problems in online learning resources.	13	8	3.77	4.00	4	.832	-.528	2	5
#12. Personally, I am skeptical about the value of automating the process of finding usability problems in online learning resources.	13	8	2.92	3.00	2	.954	.854	2	5

The Open-Ended Comments

Three out of thirteen respondents commented on the open-ended item of the first survey instrument. Their comments are as follows:

Given that the website in question was developed as part of a research project, not a development & distribution project, it is held to a somewhat different standard for usability.

My response through #7 only concerns the site which prompted my participation in this survey. That site is essentially static and was not intended to be developed further. However, current projects would benefit from a usability analysis. I have solicited it informally, and I have received limited feedback. So there is a certain “attraction” to an automated usability analysis!

Personally, usability is something I’ve been actively learning more about recently. I have some training in survey application and psychological/behavioral testing, however, I find that since our organization is small, opportunities for formal study of our site’s usability are limited. Application of methods to enhance usability are mainly done whenever I can convince the boss -- i.e. ad hoc. Logically, this would make us a good candidate for automated testing, though I remain skeptical a program could be created that effectively channels the random behaviors of the elusive “average user”.

In summary, the survey results revealed the following about the *perception* among Web-based learning site builders regarding usability evaluation:

- They *agreed* that they are interested in further improvement of the usability of their Web sites.
- They *strongly agreed* that usability is an essential feature of online learning resources.
- They were neutral (*neither agreed nor disagreed*) with respect to whether usability is usually overlooked in the design of online learning resources.

- They *agreed* that they would like to know more about usability in the context of online learning resources.
- They *agreed* that it would be very helpful to automate the process of finding usability problems in online learning resources.
- They were neutral (*neither agreed nor disagreed*) with respect to whether they are skeptical about the value of automating the process of finding usability problems in online learning resources.

Findings of Research Questions #3

As described in Chapter 1, research question #3 of this study is as follows:

3. According to the perception of Web-based learning site builders, how useful are the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality?

The second survey instrument was designed to address research question #3. This subsection describes the analysis results of participant responses to questions #1 through #4 of the second survey instrument, which address research question #3, as shown in Figure 4.5.

I am interested in your perceptions of the summary of the usability evaluation results of your Web site that you have received from me recently. Please check your agreement with each of the following statements using the scale below. 1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable						
	1	2	3	4	5	N/A
#1. The summary of the usability results regarding my Web site was easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#2. The results accurately reflect the usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#3. The results accurately reflect usability issues of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#4. The results would be helpful for improving usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Figure 4.5. Questions #1 through #4 of the second survey instrument which addressed research question #3.

Descriptive statistics and histograms were used to describe the *central tendency, variability, size, and distribution* of the responses to survey questions #1 through #4 of the second instrument (see Figure 4.5).

The Descriptive Statistics

Table 4-5 illustrates the descriptive statistics of the scores of survey questions #1 through #4 of the second survey instrument in SPSS default output format. Table 4-6 illustrates the association between each survey question and its descriptive statistics.

Figure 4.6 illustrates the mean scores of survey questions #1 through #4 of the second survey instrument in the pilot study.

Table 4-5

The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #1 through #4 of the Second Survey Instrument in the pilot study

Statistics		Q1	Q2	Q3	Q4
N	Valid	10	10	10	10
	Missing	8	8	8	8
Mean		4.10	3.60	3.60	3.30
Median		4.00	4.00	4.00	3.50
Mode		4 ^a	4	4	4
Std. Deviation		.994	.699	.699	.823
Skewness		-1.085	-1.658	-1.658	-.687
Std. Error of Skewness		.687	.687	.687	.687
Minimum		2	2	2	2
Maximum		5	4	4	4

a. Multiple modes exist. The smallest value is shown

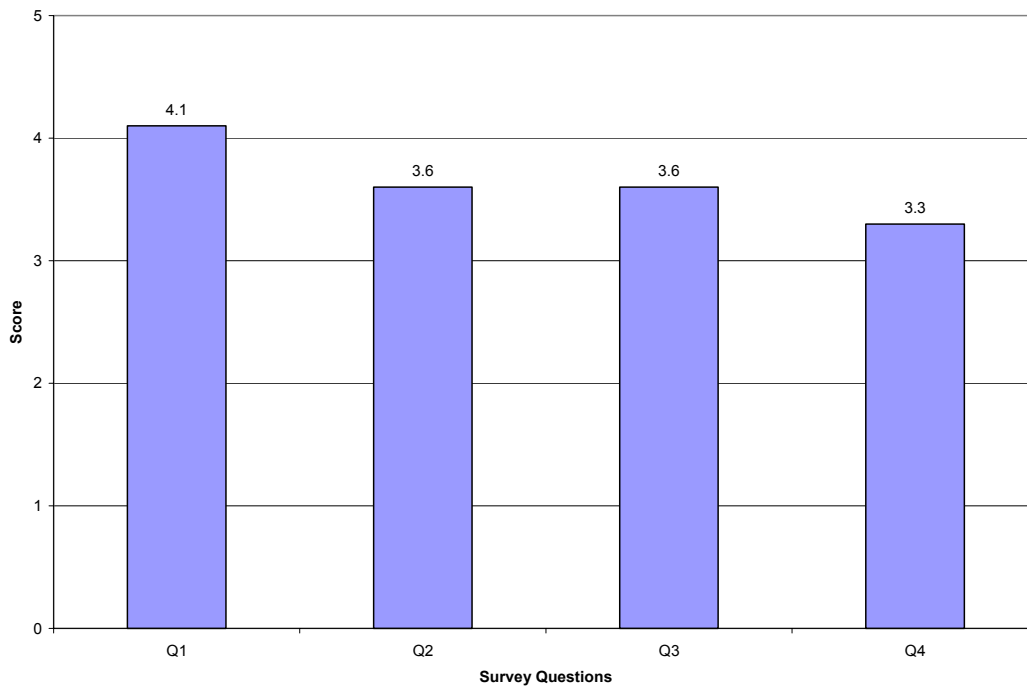


Figure 4.6. The average scores of questions #1 through #4 of the second survey instrument in the pilot study.

Table 4-6

The Descriptive Statistics of Survey Questions #1 through #4 of the Second Survey Instrument in the Pilot Study

Survey Questions	N	Missing	Mean	Median	Mode	SD	Skew	Min	Max
#1. The summary of the usability results regarding my Web site was easy to understand.	10	8	4.10	4.00	4	.994	-1.085	2	5
#2. The results accurately reflect the usability quality of my Web site.	10	8	3.60	4.00	4	.699	-1.658	2	4
#3. The results accurately reflect usability issues of my Web site.	10	8	3.60	4.00	4	.699	-1.658	2	4
#4. The results would be helpful for improving usability quality of my Web site.	10	8	3.30	3.50	4	.823	-.687	2	4

The Open-Ended Comments

Seven out of the ten surveyed responded to the open-ended item of the second survey instrument in the pilot study. Their comments are presented below.

This evaluation did not tell me anything I don't already know. It wasn't specific enough, though, to be very helpful. I.e., the explanations of the criteria describe what we are already doing, but we are still ranked "needs improvement".

Can you apply this to projects currently in development (which are not in DLESE)? Thanks for allowing me to participate.

The evaluation results are helpful, up to a point, but don't address issues such as suitability for certain age groups and usability by those with disabilities.

Like all automated usability evaluations [sic] I have seen, there is not enough context here to be helpful. It is too general with no specifics on what problems were identified and how they could be fixed.

The rating of Excellent, Good, or Room for Improvement is useful for getting an overall idea of whether the site needs significant attention, or is likely to be okay. To use the results for improving the site, I think it might be helpful to also list the particular items on the web pages that failed to meet the conditions they were tested against. This is an interesting study. Good luck, and I'd be interested in seeing the results when your dissertation is published.

I'm on the fence about the usefulness of the summary results. Perhaps for very basic analysis of the site, the automation is accurate, but I wonder about the finer points it is missing? Our website was rated highly by your program, though I know we get many messages from users that have difficulty navigating the depth of our site (we have 1000+ pages). I feel as if the automated program only touched on perhaps, the first look from a visitor, not answering questions such as "I need "A", how do I get to it from here?" I think your program accurately reflected what it measured, I am just not sure that what it measured was enough. I'm on the fence about it.

(1) Summary presented adequately, but not as clear as it could be. (2) Biggest point: the correlation between the feedback and the specific guidelines is too loose to be useful. Not sure which of the guidelines on the usability site pertain to the feedback given by the automated summary (too much effort for me to read through all the possibilities and then figure

out which were the points that triggered the feedback). (3) In general, as a designer I do not agree with all the guidelines, though I understand that guidelines are only intended to be generalized.

In summary, the survey results revealed the following about the *perception* among Web-based learning site builders regarding usefulness of the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality:

- They *agreed* that the summary of the usability results regarding their Web sites was easy to understand.
- They *agreed* that the results accurately reflect the usability quality of their Web sites.
- They *agreed* that the results accurately reflect usability issues of their Web sites.
- They were neutral (*neither agreed nor disagreed*) with respect to whether the results would be helpful for improving usability quality of their Web sites.

Summary

The survey results from the pilot test suggested that Web-based learning site builders have a relatively low level of expertise and seldom apply usability methods for developing learning Web sites. The site builders were interested in improving the usability of their Web sites, believed that usability is an essential feature of online learning, yet is often overlooked, and would like to know more about usability evaluation. They thought that using an automated tool may be helpful for identifying usability problems in online learning resources, although there was some skepticism about automated tools.

Web-based learning site builders agreed that the evaluation results provided were easy to understand, and they generally agreed that the results accurately reflected the usability quality and issues of their Web sites. They were undecided about whether the evaluation results would be helpful for improving usability quality of their Web sites.

This pilot study revealed several useful findings about conducting a survey study. The following findings were helpful in conducting the *actual study*:

- Including the Web address of the participant in the subject of an e-mail invitation increases the response rate.
- Sending multiple reminders is necessary for increasing the response rate.
- Sending an invitation for both survey instruments in one e-mail increases the chance that people will complete the second survey instrument.
- Making a follow-up phone call after sending out e-mail reminders will get people to take action. Several respondents overlooked the e-mail invitations and reminders; however, on the phone, they asked for the invitation e-mail to be resent.
- No mechanical issues related to the survey Web sites or the evaluation results Web sites were reported. However, one respondent of the actual survey reported via e-mail that the format of the survey Websites “did not render well”. I was not able to reproduce this problem.
- The data analysis procedure of the pilot study was repeated in the actual study.

Chapter 5: Results

As described in Chapter One, the two major goals of this research are as follows: to gain insight into the current state of Web-based learning sites' usability based on an automated evaluation tool, and to understand the perceptions of Web-based learning site builders about usability issues in general and about the usefulness of usability analysis results generated by an automated tool in specific.

The following research questions were addressed during this study:

1. What is the usability quality of Web-based learning sites according to an automated usability evaluation tool?
2. According to the perception of Web-based learning site builders,
 - 2.1. What is the degree of expertise among Web-based learning site builders regarding usability evaluation methods?
 - 2.2. How frequently do the Web-based learning site builders use usability evaluation methods?
 - 2.3. What is the *perception* among Web-based learning site builders regarding usability evaluation?
3. According to the perception of Web-based learning site builders, how useful are the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality?

Chapter 3 described the methodology employed for this dissertation, which was carried out in two phases (see Figure 3.1): *Phase I* consisted of a usability analysis of Web-based learning sites using WebTango and addressed research question #1. *Phase II* consisted of a survey study of Web-based learning site builders and had two survey

instruments. This phase of the study addressed research questions #2.1, #2.2, #2.3, and #3. Chapter 4 described the results of a pilot study used to refine the survey aspects of the full study. This chapter describes the results of *phase I* and *II* of this study is organized based on the research questions.

Results of Research Question #1

As described in Chapter 1, the research question #1 of this study is as follows:

1. What is the usability quality of Web-based learning sites according to an automated usability evaluation tool?

To address this research question, 130 Web-based learning sites from DLESE-DRC were evaluated using WebTango. Eight WebTango measures were included: *text element*, *text formatting*, *graphic element*, *graphic formatting*, *link element*, *link formatting*, *page title*, and *page formatting* (see Table 3-1 and Appendix B: The WebTango Web Interfaces Usability Measurements). WebTango rated each measure of a Web site (e.g., text element) as *good*, *average* or *poor*. In *phase I*, data analysis used alternative labeling and numerical coding to represent WebTango ratings (see Table 5-1)—the alternate textual labeling for increasing descriptiveness and the numerical coding for statistical analysis. The rating terms yielded by WebTango (*good*, *average*, *poor*) were changed to “*above average*, *average*, and *below average*” at the advice of my committee during a meeting held to review the results of the pilot study

Descriptive statistics and *histograms* were used to describe the *central tendency*, *variability*, *size*, and *distribution* of the ratings of eight WebTango measures of the 130 sites.

Table 5-1

Evaluation Rating Equivalence

WebTango Ratings	Equivalent Labels in this Study	Numerical Code
good	above average	3
average	average	2
poor	below average	1

The Descriptive Statistics

Table 5-2 illustrates the descriptive statistics of ratings generated by all eight WebTango measures. Figure 5.1 illustrates the frequency distribution of WebTango ratings of the eight measures of the 130 sites. WebTango did not generate results for every measure depending upon the nature of the specific site. For instance, WebTango rated all 130 sites for *link formatting* and 120 sites for *page title*, while rating only 16 sites for *page formatting*. The rules used by WebTango for excluding a site from being evaluated were not found in WebTango-related literature.

Table 5-2

The Number of Ratings (N), Mean, and Standard Deviation of the Eight WebTango Measures of the 130 Sites Included in Phase I of this Study

Statistics		TELEMENT	TFORMAT	GELEMENT	GFORMAT	LELEMENT	LFORMAT	PTITLE	PFORMAT
N	Valid	30	66	99	84	34	130	120	16
	Missing	100	64	31	46	96	0	10	114
Mean		1.4667	1.9091	1.9697	1.9762	1.5588	1.9615	1.8250	2.4375
Median		1.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.5000
Mode		1.00	1.00	1.00	1.00	2.00	1.00	1.00	3.00
Std. Deviation		.62881	.81764	.85062	.82105	.56091	.84820	.81645	.62915
Skewness		1.025	.171	.058	.045	.303	.074	.334	-.653
Std. Error of Skewness		.427	.295	.243	.263	.403	.212	.221	.564
Minimum		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum		3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00

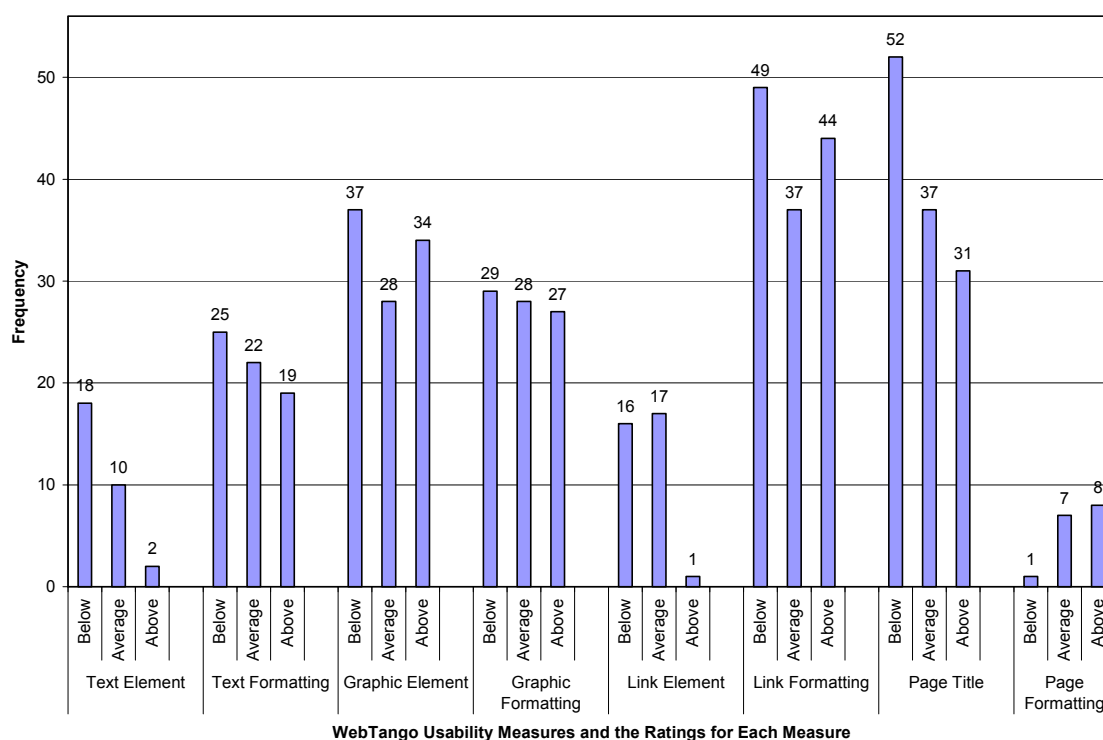


Figure 5.1. The frequency distribution of the eight WebTango measures of the 130 sites included in Phase I of this Study.

In summary, the *descriptive statistics* and *histogram* of WebTango ratings (see Table 5-1 and Figure 5.1) indicate that the learning resources rated from the DLESE-DRC were *average* on four of the measures (i.e., *text formatting*, *graphic element*, *graphic formatting*, and *link formatting*). For this sample of DLESE-DRC learning sites, three measures (i.e., *text element*, *link element*, and *page title*) were rated *below average*. Only one measure, *page formatting*, was rated *above average*; however, WebTango generated the least number of ratings (16 out of 130) for *page formatting*, and thus less confidence can be given to the results for this measure. In conclusion of the results of *Phase I* of the study, the results of the automated usability evaluation using WebTango

indicated the following: the usability quality of Web-based learning sites according to an automated usability evaluation tool is mostly *average* or *below average*.

Sites Included in the Survey Study

Thirty-five DLESE-DRC Web site builders were successfully identified and recruited to participate in the survey study. A separate analysis was conducted to determine whether the usability quality of the 35 sites belonging to these 35 participants is consistent with the usability quality of the overall sample of 130 sites. These 35 sites were a subset of the 130 sites analyzed in phase I of this dissertation, and it was deemed important to establish that this smaller subset was not significantly different from the larger sample. This subsection describes the usability quality ratings of the 35 Web sites. The methodology for analyzing these 35 sites was identical to the analysis conducted during *phase I* of this dissertation (see Results of Research Question #1 section of this chapter). *Descriptive statistics* and *histograms* were used to describe the *central tendency, variability, size, and distribution* of WebTango ratings of the eight measures of the 35 sites.

The Descriptive Statistics

Table 5-3 illustrates the descriptive statistics of the ratings generated by all eight WebTango measures of the 35 sites. WebTango did not generate results for every measure depending upon the nature of the specific site. For instance, WebTango generated *text element* ratings for only 7 out of the 35 sites, while the automated tool generated *link formatting*, and *page title* ratings for all 35 sites.

Table 5-3

The Number of Ratings (N), Mean, and Standard Deviation of the Eight WebTango Measures of 35 Sites Included in Phase II

Statistics		TELEMENT	TFORMAT	GELEMENT	GFORMAT	LELEMENT	LFORMAT	PTITLE	PFORMAT
N	Valid	11	27	32	30	9	35	35	7
	Missing	24	8	3	5	26	0	0	28
Mean		1.4545	2.0741	2.1563	2.1333	1.3333	1.9143	1.7429	2.4286
Median		1.0000	2.0000	2.0000	2.0000	1.0000	2.0000	1.0000	2.0000
Mode		1.00	3.00	3.00	3.00	1.00	1.00	1.00	2.00
Std. Deviation		.52223	.87380	.88388	.86037	.50000	.91944	.88593	.53452
Skewness		.213	-.151	-.323	-.270	.857	.177	.546	.374
Std. Error of Skewness		.661	.448	.414	.427	.717	.398	.398	.794
Minimum		1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.00
Maximum		2.00	3.00	3.00	3.00	2.00	3.00	3.00	3.00

Figure 5.2 illustrates the frequency distribution of the WebTango ratings of the eight measures for 35 sites.

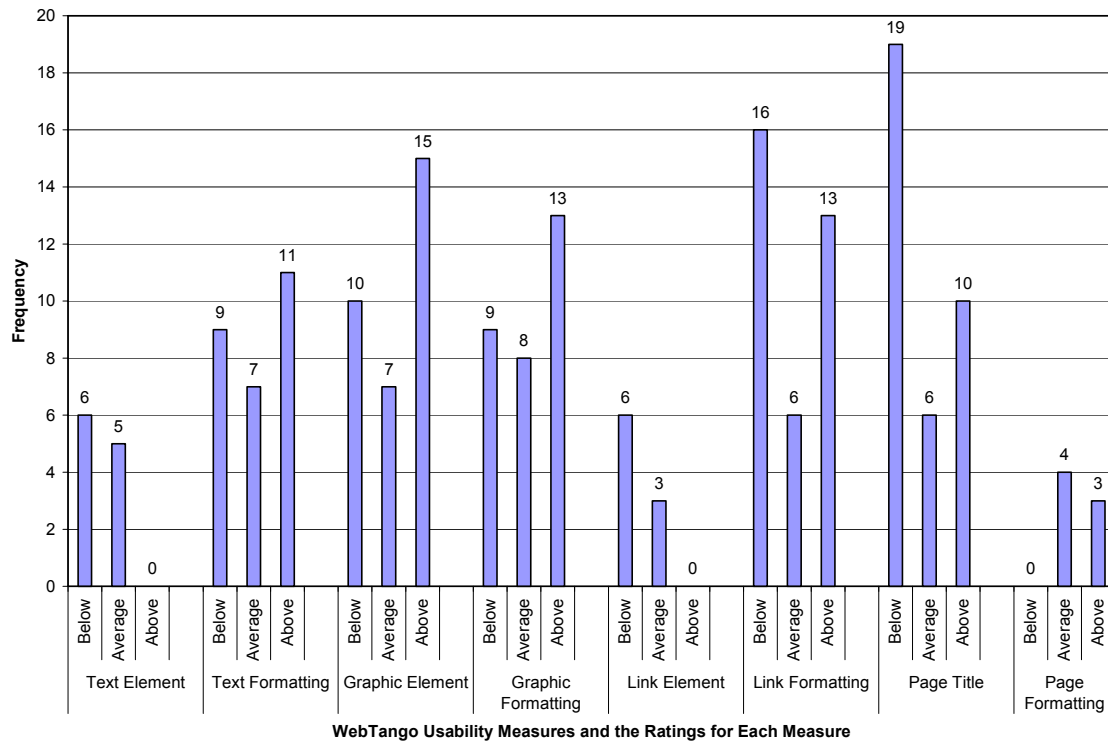


Figure 5.2. The frequency distribution of the eight WebTango measures of 35 sites.

The Results of All Sites vs. the Survey Sample Sites

This subsection describes the usability quality comparisons between these groups: *group 1*, all 130 sites and *group 2*, the 35 sites included in the survey study. Table 5-4 illustrates the descriptive statistics of the ratings generated by all eight WebTango measures of both groups. Figure 5.3 illustrates the trends of the means.

A *t-test* was used to determine whether the ratings of the 130 sites based on each WebTango measure differed from the ratings of the 35 sites. The null hypothesis was that there are not any differences. Of course, for some of the measures, the sample size was too small to make a statistically meaningful comparison. Table 5-5 illustrates the *t-test* results.

Table 5-4

The Descriptive Statistics of ratings. Group 1 being all 130 sites. Group 2 being the 35 Surveyed Sites

Group Statistics					
	Group	N	Mean	Std. Deviation	Std. Error Mean
TELEMENT	1	30	1.47	.629	.115
	2	11	1.45	.522	.157
TFORMAT	1	66	1.91	.818	.101
	2	27	2.07	.874	.168
GELEMENT	1	99	1.97	.851	.085
	2	32	2.16	.884	.156
GFORMAT	1	84	1.98	.821	.090
	2	30	2.13	.860	.157
LELEMENT	1	34	1.56	.561	.096
	2	9	1.33	.500	.167
LFORMAT	1	130	1.96	.848	.074
	2	35	1.91	.919	.155
PTITLE	1	120	1.83	.816	.075
	2	35	1.74	.886	.150
PFORMAT	1	16	2.44	.629	.157
	2	7	2.43	.535	.202

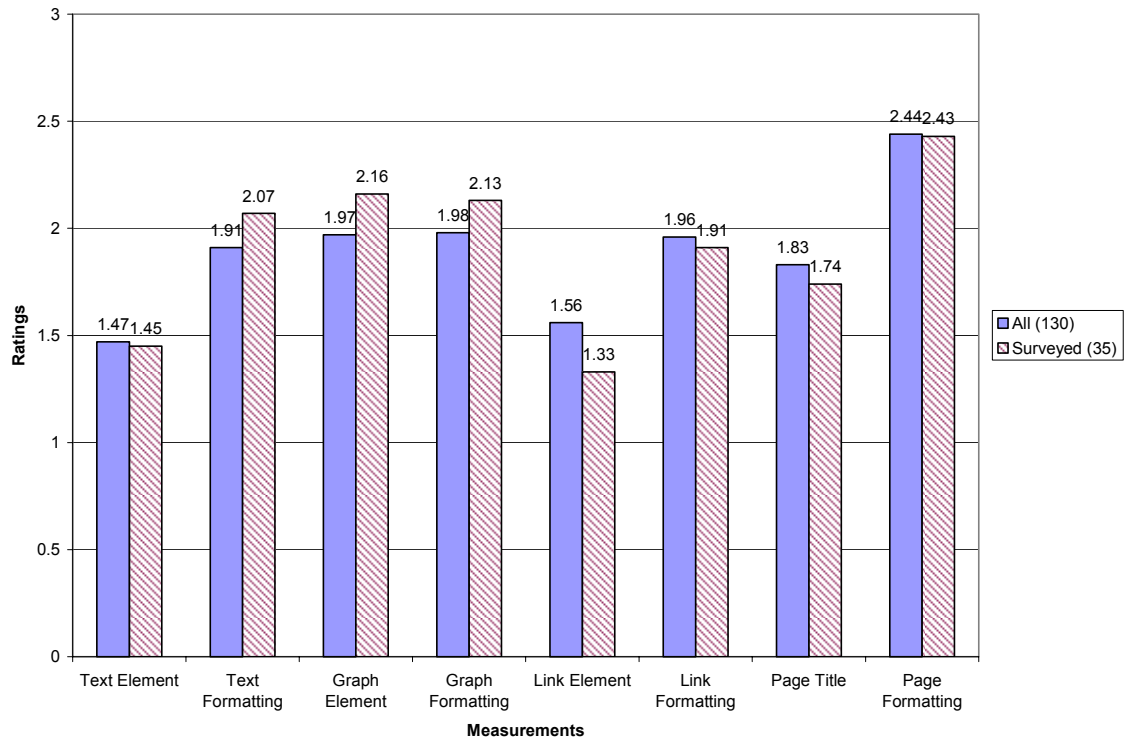


Figure 5.3 The trends of the means of the two groups: Group 1 being all 130 sites. Group 2 being the 35 surveyed sites.

The *t*-test results for each of the eight WebTango measures of *all 130 sites* and the *35 survey sample sites* based on the two-tailed hypothesis $H_0: \mu_1 - \mu_2 = 0$ at significance level $\alpha = .05$ are described below.

Text element, failing to reject H_0 : The means of *text element* of all 130 sites (1.47) and the survey sample sites (1.45) did not differ significantly [$t_{(39)} = .057$, ns]. Since the F-ratio was not significant, $F_{(1, 39)} = .620$, equal variances was assumed.

Text formatting, failing to reject H_0 : The means of *text formatting* of all 130 sites (1.91) and the survey sample sites (2.07) did not differ significantly [$t_{(91)} = -.866$, ns]. Since the F-ratio was not significant, $F_{(1, 91)} = .453$, equal variances was assumed.

Table 5-5

The t-test Results for Comparing the Two Groups

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
TELEMENT	Equal variances assumed	.620	.436	.057	39	.955	.01	.213	-.418	.442
	Equal variances not assumed			.062	21.375	.951	.01	.195	-.393	.417
TFORMAT	Equal variances assumed	.453	.503	-.866	91	.389	-.16	.191	-.543	.214
	Equal variances not assumed			-.842	45.620	.404	-.16	.196	-.560	.230
GELEMENT	Equal variances assumed	.592	.443	-1.068	129	.287	-.19	.175	-.532	.159
	Equal variances not assumed			-1.047	50.896	.300	-.19	.178	-.544	.171
GFORMAT	Equal variances assumed	.659	.419	-.889	112	.376	-.16	.177	-.508	.193
	Equal variances not assumed			-.869	49.117	.389	-.16	.181	-.521	.206
LELEMENT	Equal variances assumed	1.616	.211	1.095	41	.280	.23	.206	-.191	.642
	Equal variances not assumed			1.172	13.845	.261	.23	.192	-.188	.639
LFORMAT	Equal variances assumed	1.926	.167	.287	163	.774	.05	.164	-.277	.372
	Equal variances not assumed			.274	50.664	.785	.05	.172	-.299	.393
PTITLE	Equal variances assumed	1.586	.210	.514	153	.608	.08	.160	-.234	.398
	Equal variances not assumed			.491	52.019	.625	.08	.167	-.254	.418
PFORMAT	Equal variances assumed	.594	.449	.033	21	.974	.01	.274	-.560	.578
	Equal variances not assumed			.035	13.495	.973	.01	.256	-.542	.560

Graphic element, failing to reject H_0 : The means of *graphic element* of all 130 sites (1.97) and survey sample sites (2.16) did not differ significantly [$t_{(129)} = -1.068$, ns]. Since the F-ratio was not significant, $F_{(1, 129)} = .592$, equal variances was assumed.

Graphic formatting, failing to reject H_0 : The means of *graphic formatting* of all 130 sites (1.98) and the survey sample sites (2.13) did not differ significantly [$t_{(112)} = -.889$, ns]. Since the F-ratio was not significant, $F_{(1, 112)} = .659$, equal variances was assumed.

Link element, failing to reject H_0 : The means of *link element* of all 130 sites (1.56) and the survey sample sites (1.33) did not differ significantly [$t_{(41)} = 1.095$, ns]. Since the F-ratio was not significant, $F_{(1, 41)} = 1.616$, equal variances was assumed.

Link formatting, failing to reject H_0 : The means of *link formatting* of all 130 sites (1.96) and the survey sample sites (1.91) did not differ significantly [$t_{(163)} = .287$, ns]. Since the F-ratio was not significant, $F_{(1, 163)} = 1.926$, equal variances was assumed.

Page title, failing to reject H_0 : The means of *page title* of all 130 sites (1.83) and the survey sample sites (1.74) did not differ significantly [$t_{(153)} = .514$, ns]. Since the F-ratio was not significant, $F_{(1, 153)} = 1.586$, equal variances was assumed.

Page formatting, failing to reject H_0 : The means of *page formatting* of all 130 sites (2.44) and the survey sample sites (2.43) did not differ significantly [$t_{(21)} = .033$, ns]. Since the F-ratio was not significant, $F_{(1, 21)} = .594$, equal variances was assumed.

In summary, the *descriptive statistics* and *histogram* of WebTango ratings (see Table 5-3) indicated that most of the WebTango ratings of the 35 sites were *average*, with two *below average*. The results, similar to the results of the 130-site study, indicated the following: the usability quality of Web-based learning sites, according to an automated usability evaluation tool, is mostly *average*, with some *below average*.

Results of Research Questions #2.1, and #2.2

As described in Chapter 1, and previously in this chapter, research questions #2.1 and #2.2 are as follows:

2. According to the perception of Web-based learning site builders,
 - 2.1. What is the degree of *expertise* among Web-based learning site builders regarding usability evaluation methods?

2.2. How frequently do the Web-based learning site builders *use* usability evaluation methods?

The first survey instrument was used to address questions #2.1, #2.2, and #2.3.

This subsection describes the analysis results of participant responses to questions #1 through #6 of the first survey instrument, which address research questions #2.1 and #2.2, as shown in Figure 5.4.

Please check your **level of expertise** with the following [usability examination methods](#) using a scale where 1 = "Not familiar with it" to 5 = "expert in its application."

#1. Usability Testing: User(s) perform predetermined tasks while being observed and recorded.

#2. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.

#3. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.

Please check your **use** of the following [usability examination methods](#) using a scale where 1 = "Never" to 5 = "always."

#4. Usability Testing: User(s) perform predetermined tasks while being observed and recorded.

#5. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.

#6. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.

Figure 5.4. Survey questions #1 through #6 that address research questions 2.1 and 2.2.

Descriptive statistics and histograms were used to describe the *central tendency, variability, size, and distribution* of the responses to survey questions #1 through #6 of the first instrument (see Figure 5.4).

The Descriptive Statistics

Table 5-6 illustrates the descriptive statistics of the scores of survey questions #1 through #6 of the first survey instrument in SPSS default output format. Table 5-7 illustrates the association between each survey question and its descriptive statistics.

Table 5-6

The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #1 through #6 of the First Survey Instrument

Statistics		Q1	Q2	Q3	Q4	Q5	Q6
N	Valid	30	30	28	29	29	27
	Missing	0	0	2	1	1	3
Mean		2.57	2.53	1.86	2.24	2.34	1.70
Median		3.00	2.50	1.50	2.00	2.00	1.00
Mode		3 ^a	3	1	1	1	1
Std. Deviation		1.135	1.279	1.079	1.272	1.446	1.137
Skewness		-.101	.450	1.256	.741	.637	1.655
Std. Error of Skewness		.427	.427	.441	.434	.434	.448
Minimum		1	1	1	1	1	1
Maximum		4	5	5	5	5	5

a. Multiple modes exist. The smallest value is shown

Figure 5.5 illustrates the mean scores of survey questions #1 through #6 of the first survey instrument. Figure 5.7 illustrates the mean scores comparisons between the actual survey and the pilot survey of survey questions #1 through #6 of the first survey instrument.

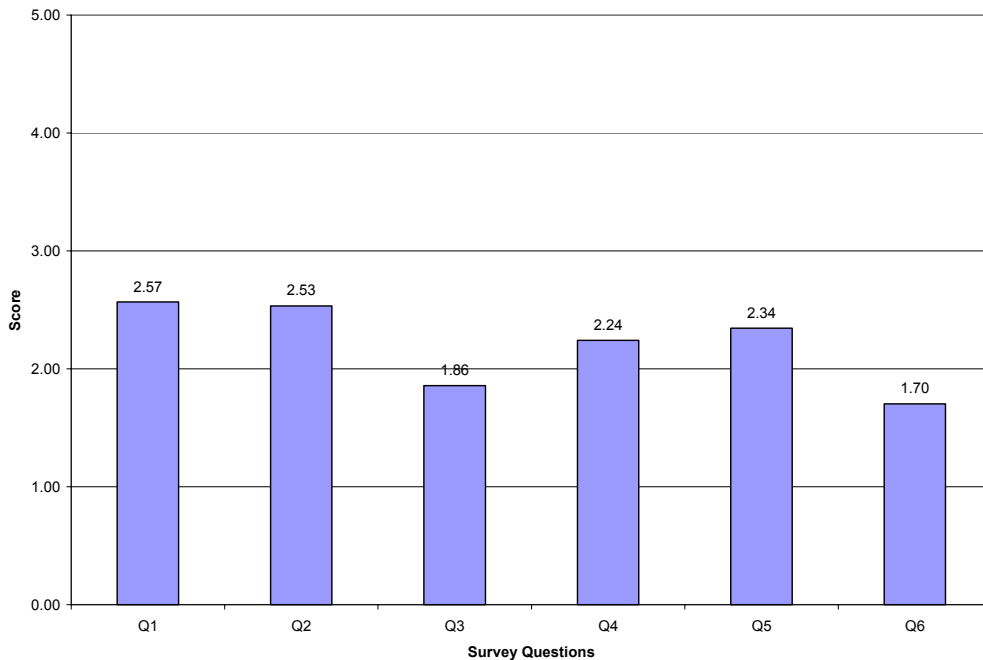


Figure 5.5. The average expertise and use of usability methods.

Table 5-7

The Descriptive Statistics of Survey Questions #1 through #6 of the First Survey Instrument

Survey Questions	N	Missing	Mean	Median	Mode	SD	Skew	Min	Max
#1. Usability Testing: User(s) perform pre-determined tasks while being observed and recorded.	30	0	2.57	3.00	3	1.135	-.101	1	4
#2. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.	30	0	2.53	2.50	3	1.279	.450	1	5
#3. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.	28	2	1.86	1.50	1	1.079	.1256	1	5
#4. Usability Testing: User(s) perform pre-determined tasks while being observed and recorded.	29	1	2.24	2.00	1	1.272	.741	1	5
#5. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.	29	1	2.34	2.00	1	1.446	.637	1	5
#6. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.	27	3	1.70	1.00	1	1.137	.1655	1	5

In summary, the survey results revealed that the mean scores of survey questions #1 through #6 of the first survey instrument (on a scale of 1 (not familiar with it) to 5 (expert in its application)), the level of expertise in usability methods of the survey participants was 2 (not very familiar with it). On a scale of 1 (not familiar with it) to 5 (expert in its application), the level of expertise in usability methods of the survey participants was 2 (not very familiar with it). On a scale of 1 (never) to 5 (always), the participants almost never use usability methods.

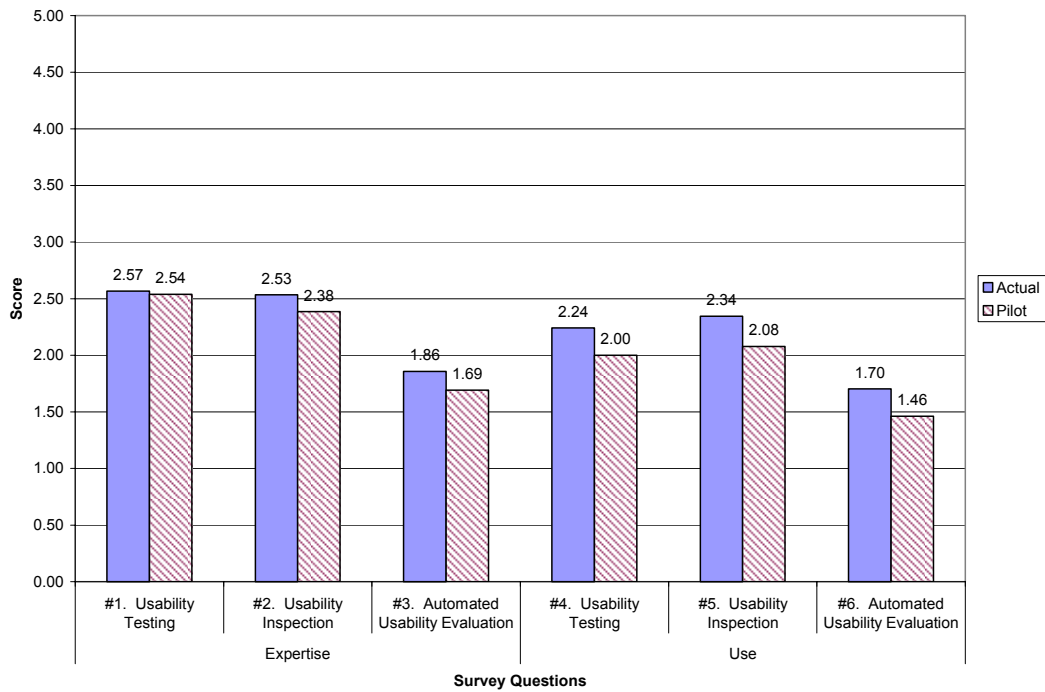


Figure 5.6 The mean scores (of questions #1 through #6) comparisons between the actual survey and the pilot survey.

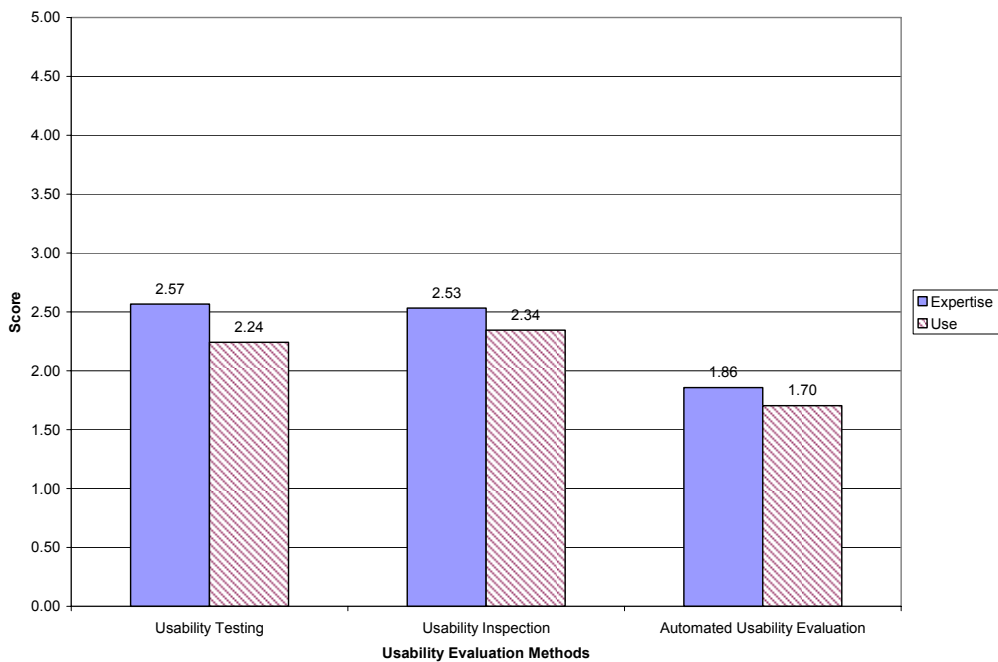


Figure 5.7 The average expertise and use of usability methods comparisons between the actual survey and the pilot survey.

These results indicated the following:

- The degree of expertise among Web-based learning site builders regarding usability evaluation was relatively low.
- Web-based learning site builders used usability evaluation methods infrequently.

Results of Research Questions #2.3

As described in Chapter 1, another research question addressed in this study is as follows:

2. According to the perception of Web-based learning site builders,

2.3 What is the *perception* among Web-based learning site builders regarding usability evaluation?

This subsection describes the analysis results of participant responses to questions #7 through #12 of the first survey instrument, which address research question #2.3, as shown in Figure 5.8.

Descriptive statistics and histograms were used to describe the *central tendency, variability, size, and distribution* of the responses to survey questions #7 through #12 of the first instrument (see Figure 5.8).

Please check your agreement with each of the following statements using the scale below.
1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

#7. I am interested in further improvement of the usability of my Web site.

#8. Usability is an essential feature of online learning resources.

#9. Usability is usually overlooked in the design of online learning resources.

#10. Personally, I would like to know more about usability in the context of online learning resources.

#11. It would be very helpful to automate the process of finding usability problems in online learning resources.

#12. Personally, I am skeptical about the value of automating the process of finding usability problems in online learning resources.

Figure 5.8. Questions #7 through #12 of the first survey instrument which addressed research question 2.3.

The Descriptive Statistics

Table 5-8 illustrates the descriptive statistics of the scores of survey questions #7 through #12 of the first survey instrument in SPSS default output format. Table 5-9 illustrates the association between each survey question and its descriptive statistics.

Table 5-8

The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #7 through #12 of the First Survey Instrument

Statistics		Q7	Q8	Q9	Q10	Q11	Q12
N	Valid	30	30	30	30	30	29
	Missing	0	0	0	0	0	1
Mean		4.53	4.67	3.27	4.13	3.93	2.90
Median		5.00	5.00	3.00	4.00	4.00	3.00
Mode		5	5	3	4	3	3
Std. Deviation		.629	.547	1.202	.730	.868	.900
Skewness		-1.025	-1.407	-.300	-.214	.134	-.732
Std. Error of Skewness		.427	.427	.427	.427	.427	.434
Minimum		3	3	1	3	3	1
Maximum		5	5	5	5	5	4

Figure 5.9 illustrates the mean scores of survey questions #7 through #12 of the first survey instrument.

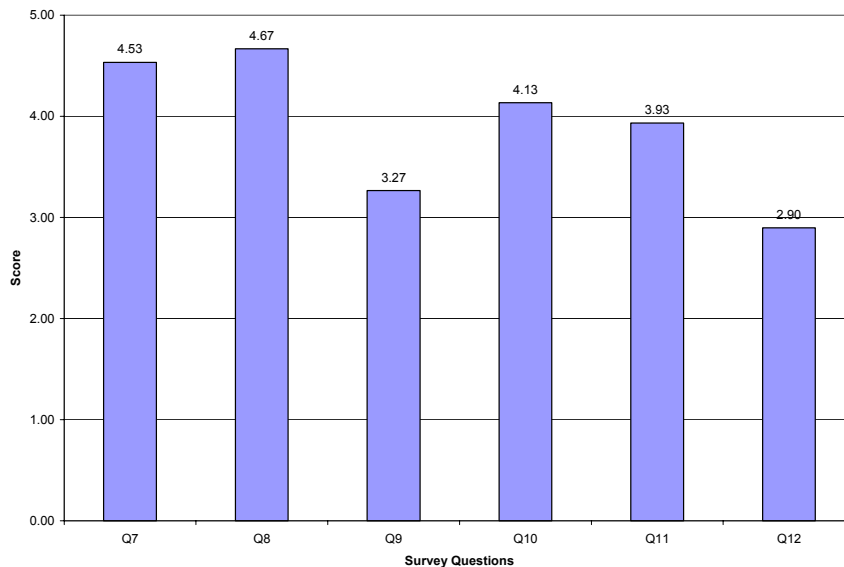


Figure 5.9. The average scores of survey questions #7 - #12 of the first survey instrument.

Table 5-9

The Descriptive Statistics of Survey Questions #7 through #12 of the First Survey Instrument

Survey Questions	N	Missing	Mean	Median	Mode	SD	Skew	Min	Max
#7. I am interested in further improvement of the usability of my Web site.	30	0	4.53	5.00	5	.629	-1.025	3	5
#8. Usability is an essential feature of online learning resources.	30	0	4.67	5.00	5	.547	-1.407	3	5
#9. Usability is usually overlooked in the design of online learning resources.	30	0	3.27	3.00	3	1.202	-.300	1	5
#10. Personally, I would like to know more about usability in the context of online learning resources.	30	0	4.13	4.00	4	.730	-.214	3	5
#11. It would be very helpful to automate the process of finding usability problems in online learning resources.	30	0	3.93	4.00	3	.868	.134	3	5
#12. Personally, I am skeptical about the value of automating the process of finding usability problems in online learning resources.	29	1	2.90	3.00	3	.900	-.732	1	4

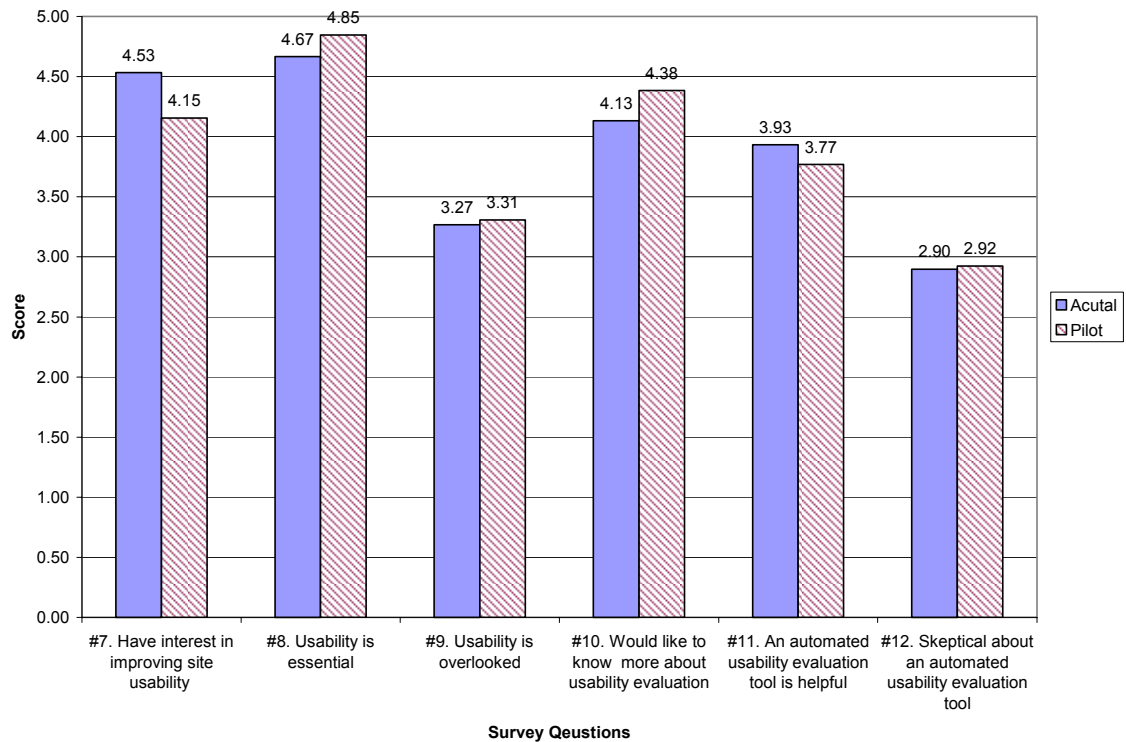


Figure 5.10 The mean scores (of questions #7 through #12) comparisons between the actual survey and the pilot survey.

The Open-Ended Comments

Ten of the 30 participants responded to the open-ended item of the first survey instrument. Their comments are listed below with their original spelling and grammar errors retained.

Do usability surveys discriminate between the knowledgeable and non-knowledgeable users, and are separate usability metrics calculated?

I think this study is about determining (1) ease of navigation through the website from page to page to find information (2) page layout (3) ease of finding information (4) clarity of information presentation. Is this what is meant by usability?

I believe in usability testing and will use any methodology that will help to improve my websites. However, usability testing, inspection or automated evaluation are only as good as the criteria used. Even user testing is only as good as the set of predetermined tasks. Your survey omits a highly relevant test (the “acid test”) of website usability, which is the Google ranking and the number of hits compared to similar websites. I’ve been coordinating development of science websites since the early 1990’s, and my websites do rank high in Google (always on the first results page, often first or in the top five). My science websites have also been reviewed favorably by prestigious sources such as Science magazine and Scientific American magazine. Over the years, I have developed my own criteria based on user feedback in the form of email to the webmaster, and some usability efforts of my own making. I have given invited presentations on how to build successful websites.... I have never used automated tools, but I am interested, and would try them. I value additional usability testing because it may help me to make my websites even more useful, or useful to a broader audience. However, I am pragmatic: any usability test should be evaluated on the basis of its track record in improving the Google ranking or increasing the hits on the website, or in favorable reviews, once the usability issues have been addressed. I hope that your thesis on investigating usability will include followup to see if your usability tests lead to improvements in website usage by some independent measure, such as number of hits or Google ranking.

Questions are directed to general usability in this survey. We are also looking into different levels of usability. We go beyond 508 compliance and now test for XHTML and CSS compliance.

I usually go with gut reaction. I never heard of software that performs this task.

I have used Bobby in the past to check for usability and typically find so many things that appear to be “false-positives”. I am skeptical about automated tools.

The arena of online education is so new, any help we can offer one another will go a long way towards improving usability issues, certainly!

I would be interested in learning more about the potential of using automated processes (software, etc.) to find usability problems.

I did not do repetative [sic] usability tests on this gw gip [sic] site, but have done so on other sites since this one.

This is critical for teachers to learn, not just students

In summary, the survey results revealed the following about the *perception* among Web-based learning site builders regarding usability evaluation:

- They *strongly agreed* that they are interested in further improvement of the usability of their Web sites.
- They *strongly agreed* that usability is an essential feature of online learning resources.
- They *agreed* that usability is usually overlooked in the design of online learning resources.
- They *agreed* that they would like to know more about usability in the context of online learning resources.
- They *agreed* that it would be very helpful to automate the process of finding usability problems in online learning resources.

- They were neutral (*neither agreed nor disagreed*) with respect to whether they are skeptical about the value of automating the process of finding usability problems in online learning resources.

Findings of Research Question #3

As described in Chapter 1, research question #3 of this study is as follows:

3. According to the perception of Web-based learning site builders, how useful are the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality?

The second survey instrument was designed to address research question #3. This subsection describes the analysis results of participant responses to questions #1 through #4 of the second survey instrument, which address research question #3, as shown in Figure 5.11.

I am interested in your perceptions of the summary of the usability evaluation results of your Web site that you have received from me recently. Please check your agreement with each of the following statements using the scale below.
1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

	1	2	3	4	5	N/A
#1. The summary of the usability results regarding my Web site was easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#2. The results accurately reflect the usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#3. The results accurately reflect usability issues of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#4. The results would be helpful for improving usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Figure 5.11. Questions #1 through #4 of the second survey instrument which address research question #3.

Descriptive statistics and histograms were used to describe the *central tendency*, *variability*, *size*, and *distribution* of the responses to survey questions #1 through #4 of the second instrument (see Figure 5.11).

The Descriptive Statistics

Table 5-10 illustrates the descriptive statistics of the scores of survey questions #1 through #4 of the second survey instrument in SPSS default output format. Table 5-11 illustrates the association between each survey question and its descriptive statistics.

Table 5-10

The Descriptive Statistics Generated by SPSS: Number of Scores (N), Mean, and Standard Deviation of survey Questions #1 through #4 of the Second Survey Instrument

Statistics		Q1	Q2	Q3	Q4
N	Valid	29	29	29	29
	Missing	0	0	0	0
Mean		4.00	3.52	3.31	3.38
Median		4.00	3.00	3.00	4.00
Mode		4	3	4	4
Std. Deviation		.886	.785	.891	1.015
Skewness		-1.322	.177	-.029	-.418
Std. Error of Skewness		.434	.434	.434	.434
Minimum		1	2	2	1
Maximum		5	5	5	5

Figure 5.12 illustrates the mean scores of survey questions #1 through #4 of the second survey instrument.

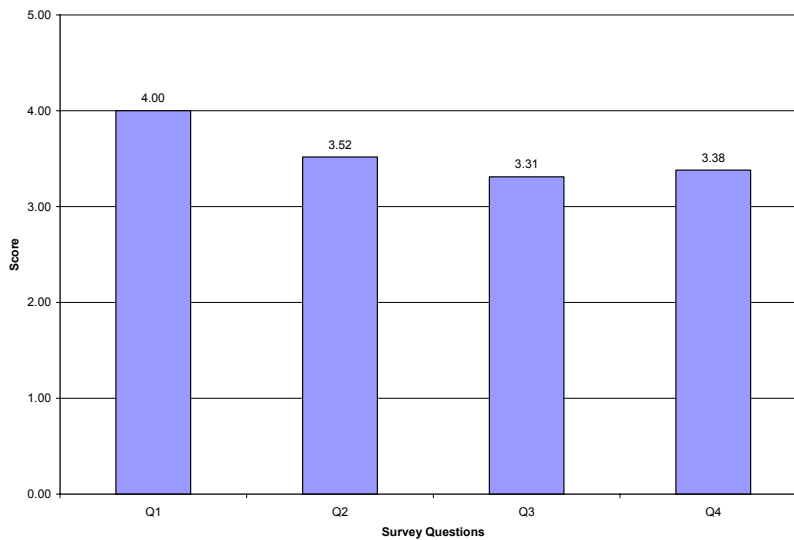


Figure 5.12. The average scores of questions #1 through #4 of the second survey instrument.

Table 5-11

The Descriptive Statistics of Survey Questions #1 through #4 of the Second Survey Instrument

Survey Questions	N	Missing	Mean	Median	Mode	SD	Skew	Min	Max
#1. The summary of the usability results regarding my Web site was easy to understand.	29	0	4.00	4.00	4	.886	-1.322	1	5
#2. The results accurately reflect the usability quality of my Web site.	29	0	3.52	3.00	3	.785	.177	2	5
#3. The results accurately reflect usability issues of my Web site.	29	0	3.31	3.00	4	.891	-.029	2	5
#4. The results would be helpful for improving usability quality of my Web site.	29	0	3.38	4.00	4	1.015	-.418	1	5

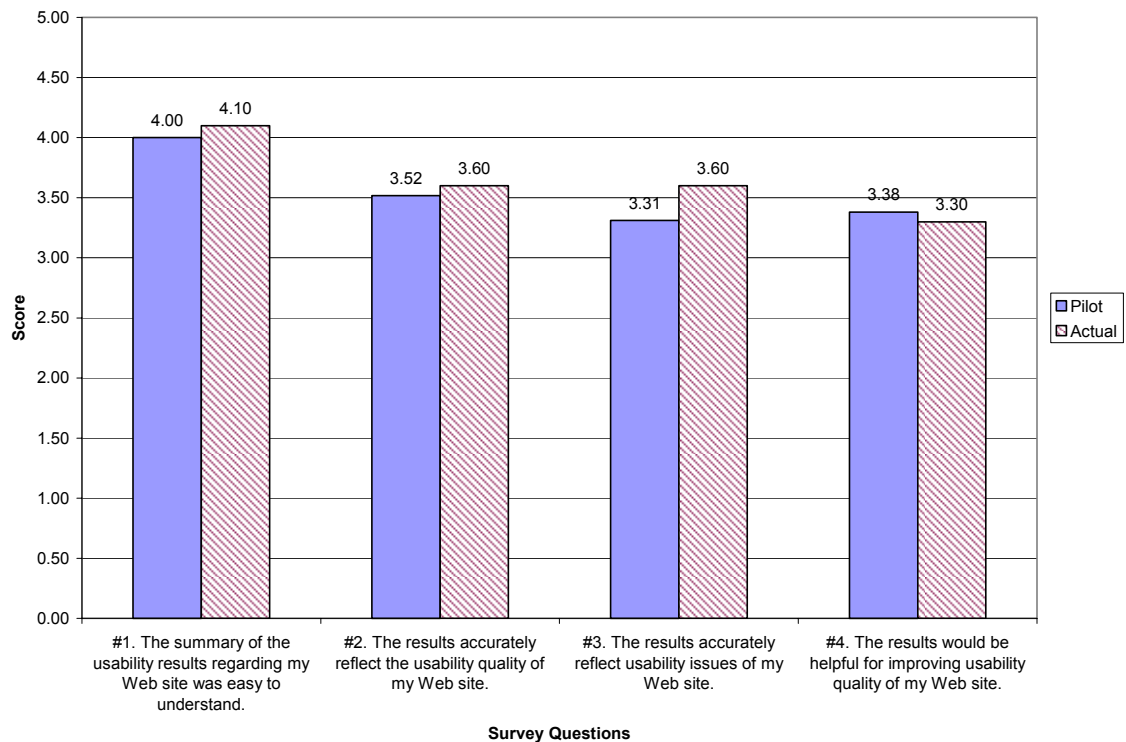


Figure 5.13 The mean scores (of questions #1 through #4) comparisons between the actually survey and the pilot survey.

The Open-Ended Comments

Eighteen of the 29 participants who completed the second survey responded to the open-ended item of the second survey instrument. Their comments are listed below with their original spelling and grammar errors retained.

I don't understand why some items are not rated.

Most of the evaluation criteria were "Not Rated". I don't believe you can evaluate the overall [sic] quality of a site based on a few minimal [sic] checks. Since most of my results were "not rated", it is hard to evaluate how helpful the ratings could be. It is helpful to know that I should go look at how the page title and links are formatted, as perhaps I could make those more user-friendly.

Rather than a one-word statement, such as "Good", it would be more beneficial to have quantified data so that I could determine what was good and what wasn't. The current system lets me know if I'm in the ball park or not, but doesn't tell me what I need to do to hit a home run. The comments lack specificity in terms of what is viewed as inadequate and what could be done within the denoted categories. I cannot use these results to make specific definitive changes to the website.

This is probably a semantic point, but the use of the term "usability" seems to me to be somewhat misleading. The automated tool seems to perform an analysis of page layout relative to several defined measures (text size, image density, etc.). I agree that the tool provides useful information from this perspective and should help in page design. The tool doesn't give any information as to how "usable" the site is in terms of content access and extraction however - I guess I was expecting some kind of measures of how many links had to be followed to extract a particular piece of major information, how easy it was to navigate through the site, etc. This kind of measure is more effectively (or perhaps needs to be) performed by a heuristic manual evaluation.

The Usability Evaluation Results would have been much more helpful to me if the full summary had been provided. The "Reading the Evaluation Results" page at <http://www.web4success.org/survey/results/textHelp.htm> contained far more information in the examples at the bottom of the page <http://www.web4success.org/survey/results/textHelp.gif>, than were provided to me in the summary. Although your text descriptions were quite clearly worded, greater specificity would have enabled me to make improvements. This interesting survey showed me some improvements that could be made, but it also showed me just how difficult it is to build

an automated tool to test usability [sic]. Perhaps this would be more useful for a less experienced web developer. There are some truly awful web pages out there.

For those with little usability background, the test results are useful for analysis and can help pinpoint specific usability items. The specific test results I received touched on the most basic usability principles [sic], not the interaction of these basic elements, how they all work together, to facilitate usability. If test results are limited to usability items as they are presented “out-of-context” this may not be a highly beneficial tool for usability professionals.

If this is a service that we could take advantage of please let me know.

While I think it is helpful to have a website reviewed by some type of automated usability evaluation, which I interpret as having a clear/static set of guidelines or standards to follow, I feel that, in regard to providing feedback, the automated system will fall short. I could see an automated system as being an initial step to a comprehensive usability evaluation. For example, do all the pages have titles, is the navigation consistent, etc. but because websites vary so drastically in content, visual presentation and so forth, and that they are designed for human interaction, I feel that there needs to be a human evaluation as well as an automated evaluation.

[Our] web site can certainly use improvement, and the usability [sic] tool probably helps to focus on some areas that are especially weak. It was not clear to me if the evidence indicators (the scale of 5 circles) applied to [our] site directly? Also, I would like to see the full web site of the helpful hints and tips for designing usability [sic] (it may be there and I missed it - that was the most useful feature that I found)

Saun, Thanks for the opportunity to participate. People pay good money for evaluations of their tools and resources! This was a good project.

While [sic] it seems that automated results are too generic, it does draw attention to pieces [sic] and parts of the webpage that might need re-visiting.

The repetitive [sic] and vague term “room for improvement” applied to every element of this site is not helpful, and makes me wonder if automated usability results are always identical? The link to specific standards is helpful.

The links' info that define the categories in the column on the right is more important than the redundant verbiage in the table. This topic has great promise, but I think that it is difficult and not relevant [sic] for a project to have an opinion on the results....better that the results be verified

and just be certain to be understood....that way the data can affect change.
--very interesting

The scanner appears to be too general (“you have sufficient images” “your average image dimensions are good”) to be significantly useful outside of very narrow areas

For the areas “need improvement” I neither agree or disagree [sic] as it is important to understand from the person submitting the evaluation the basis of the comments. I appreciate [sic] the effort and look forward to following up on the results. well organized effort. Thanks

It was nice to have links to usability.gov next to each category. However, I do not feel I know the specific instances on my site where usability could be improved (i.e., if I wanted to improve my “text element” usability - are there certain pages where this is particularly bad - or should I reevaluate it site-wide?) Also, it would be nice to know if users are able to find what they are looking for when they come to the site and I still do not know the answer to that. Thanks for evaluating my site.

Not enough information. What are the standards for evaluation?

If this is a service that we could take advantage of please let me know.

The comments lack specificity in terms of what is viewed as inadequate and what could be done within the denoted categories. I cannot use these results to make specific definitive changes to the website.

In summary, the results from the second survey revealed the following about the *perception* among Web-based learning site builders regarding the usefulness of the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality:

- They *agreed* that the summary of the usability results regarding their Web sites was easy to understand.
- They *agreed* that the results accurately reflected the usability quality of their Web sites.

- They were neutral (*neither agreed nor disagreed*) regarding whether the results accurately reflected usability issues of their Web sites.
- They were neutral (*neither agreed nor disagreed*) regarding whether the results would be helpful for improving usability quality of their Web sites.

Survey Response Rates

Thirty-five potential participants were invited to participate in the survey study consisting of two survey instruments. The following list illustrates the high-level steps taken in conducting the Web-based survey:

- The 35 potential participants were invited via e-mail responded to the first survey instrument (see Appendix E: Survey Instruments).
- A participant was invited via e-mail to respond to the second survey instrument when he/she had completed the first survey instrument.
- After three days, an e-mail reminder was sent to those who did not respond to the survey.
- After 50% of the survey was completed, a short reminder was sent to the remaining people. This reminder contained the URL of the first survey instrument and the URL of the usability evaluation results for each Web site, which contained the URL of the second survey instrument.
- Follow-up phone calls were made to those who did not respond to any of the prior e-mail communication. A few people indicated that they did not receive the e-mail. We concluded that the e-mails were filtered as junk mail.

The first survey instrument had 30 respondents (85.71%). The second survey instrument had 29 respondents (82.86%). Most people were very supportive of the study;

however, they were extremely busy. One person declined because he had other priorities. Two people did not respond to the e-mails or voice mail, and two others agreed to participate but did not take the survey.

Limitations

The *results* phase of the study has the following limitations: First, only 130 out of nearly 600 DLESE-DRC Web sites were analyzed. Given that an automated usability evaluation tool is capable of analyzing a large number of sites, including all DLESE-DRC Web sites was technically feasible. However, there were obstacles preventing the inclusion of all DLESE-DRC sites which were described in The Target Web Sites section of Chapter Three. Second, as described earlier, WebTango did not generate results for every measure depending upon the nature of the specific site. For instance, WebTango rated all 130 sites for link formatting and 120 sites for page title, while rating only 16 sites for page formatting. Third, only ten of the 30 participants responded to the open-ended item of the first survey instrument. In contrast, 20 of the 29 participants who completed the second survey responded to the open-ended item of the second survey instrument. This study would have contributed to a fuller understanding of the perceptions of Web site builders if a richer data set were collected. Finally, the usability evaluation report is not directly generated by WebTango. Unfortunately, the usability evaluated results generated by WebTango were very cryptic. Therefore, showing the WebTango results directly to the survey participants was not a feasible option for this study.

Summary

Based on the eight WebTango usability evaluation measures (i.e., *text element*, *text formatting*, *graphic element*, *graphic formatting*, *link element*, *link formatting*, *page title*, and *page formatting*), the current state of Web-based learning sites' usability quality is mostly *average* or *below average*.

The survey results indicated that Web-based learning site builders have a relatively low degree of expertise in usability methods and seldom apply usability methods for designing and building Web-based learning sites. Although Web-based learning site builders agreed that the WebTango ratings accurately reflected the usability quality of their Web sites, they were not sure whether the ratings accurately reflected the usability issues of their Web sites. Web-based learning sites builders were undecided about whether WebTango evaluation results would be helpful for improving usability quality of their Web sites.

Chapter 6: Conclusions and Discussion

As detailed in the earlier chapters, this dissertation has been carried out with the following two major goals: 1) to gain insight into the current state of Web-based learning site usability based on an automated evaluation tool and 2) to understand the perceptions of Web-based learning site builders about usability evaluation in general and about the usefulness of usability analysis results generated by an automated tool in specific. This last chapter includes a summary of the study, as well as conclusions, implications, and recommended future research directions.

Study Overview

The Web has become an increasingly integral part of teaching and learning in higher education since its beginning a dozen years ago (Khan, 1997). Pedagogy is the principal consideration when building a Web-based learning environment, but usability is another critically important issue. Although many studies have been conducted regarding the pedagogical aspects of Web-based learning in higher education (Bruning, Horn, & PytlikZillig, 2003; Jonassen, 2003), usability is a less evident topic in the educational research literature, and the use of automated usability evaluation tools is especially unexplored.

Web site usability has many distinctions from *software usability*, and is considered a sub-area within the term *usability*. Web-based learning usability is especially important because a site that is not usable will not support learning even if the pedagogical design is strong. According to usability experts such as Jakob Nielsen (2000a), Web sites and Web-based learning sites generally have low usability quality. Since conducting a formal usability test can be time and resource intensive, alternative

usability evaluation methods such as heuristic evaluation and automated usability evaluation have been used. Automated usability evaluation has helped Web site builders to improve usability quality, especially for those who do not have formal training in usability design and evaluation (Ivory, 2003; Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004). But automated tools have not been widely applied to online educational resources.

WebTango was used to analyze Web-based learning sites and the developers of these sites were surveyed regarding their perceptions of usability issues in general and the analysis results of their sites provided by the automated tool in specific. WebTango was selected for several reasons, the primary reason being that it is an automated evaluation tool that is based on empirical measures. It also offers a more extensive set of measures specific to usability than any other guideline review tool available today, and heuristic evaluation is an accepted method to supplement formal usability testing. WebTango is a *guideline review* usability evaluation tool intended for *information-centric* Web sites and so is suitable for many educational Web sites (Ivory, 2001).

This dissertation was conducted in two phases. In Phase I of the study, 130 of the Web-based learning sites linked to DLESE, a digital library, were evaluated using WebTango. Specifically, the study was targeted on sites included in the DLESE Reviewed Collection (DRC). To be part of the DRC, these sites had been evaluated and found to meet seven evaluation criteria including pedagogical effectiveness and ease of use. In *Phase II*, a survey study was conducted to gather the perceptions of Web-site builders about the usefulness of the WebTango analysis results in determining the usability of their own Web sites.

Conclusions

The following conclusions are made regarding the automated analysis of the Web-based learning sites and the perceptions of the site builders investigated in addressing the goals of this study:

- Despite being included in the DRC, a sample of 130 Web-based learning sites was indicated as having average or low usability quality.
- The builders of these Web-based learning sites possess a low degree of expertise in usability evaluation methods.
- The builders of these Web-based learning sites seldom apply usability evaluation methods when developing online learning resources.
- The builders of these Web-based learning sites perceive usability evaluation as important in developing online learning resources.
- The builders of these Web-based learning sites perceive that the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality is useful to a limited degree.

The following subsections discuss each of these conclusions.

Web-based Learning Sites Have Average or Low Usability Quality

The ratings of usability quality of Web-based learning sites are *average* or *below average*, according to WebTango—an automated usability evaluation tool. Four out of eight measures (*text formatting*, *graphic element*, *graphic formatting*, and *link formatting*) are *average*. Three out of eight measures (*text element*, *link element*, *page title*) are *below average*. Only one measure (*page formatting*) is *above average*. However, WebTango rated *page formatting* for only 16 of the 130 sites.

This study's findings, *average* or *below average* usability quality of Web-based learning sites, are consistent with the other indications in educational Web site-related literature (Kirschner, Strijbos, Kreijns, & Beers, 2004; Norman & Draper, 1986; Storey, Phillips, Maczewski, & Wang, 2002). Given that Web-based learning site builders have relatively low degrees of expertise and seldom apply usability evolution methods, the findings of this and other studies are not surprising.

Site Builders have Low Degree of Expertise and Seldom Apply of Usability Evaluation Methods

Web-based learning site builders have low degree of expertise of usability evaluation methods. The Web-based learning site builders' degree of expertise in usability evaluation methods is relatively *low*. On a scale of 1 to 5 (where 1 is *not familiar with it* and 5 is *expert in its application*), the survey results indicate that Web-based learning site builders' degree of expertise in usability evaluation methods is 2.

There are differences among the three types of response patterns related to the *expertise* and *use* of usability evaluation methods (see Figure 5.7). The respondents are most familiar with usability testing which is the most fundamental approach to usability evaluation. Expectedly, they are somewhat less familiar with usability inspection or heuristic evaluation, and least familiar with automated approaches.

Web-based learning site builders seldom apply usability evaluation methods when developing Web sites. The Web-based learning site builders *almost never* apply usability evaluation methods when developing a Web site. On a scale of 1 to 5 (where 1 is *never* and 5 is *always*), the survey results indicate that Web-based learning site builders score a 2 on related survey items.

With respect to *use* (see Figure 5.7), they are slightly more likely to have used usability inspection or heuristic evaluation than usability testing which is not surprising given the greater time, resource and expense required by usability testing. They are least likely to have used automated approaches; again not surprising given the relative novelty of automated usability evaluation.

Some indications in the open-end comments suggest that practical constraints (e.g., time and resources) are possible causes of this *low degree of expertise* and *infrequent use* trend. Similar constraints have prevented usability evaluation and other forms of formative evaluation from being widely applied to the design and development of educational material (Flagg, 1990). The low degree of expertise and infrequent application of usability evaluation methods among Web-based learning site builders is clearly not caused by the lack of interest among Web-based site builders. It may be that while these learning resource builders care about usability, they pay greater attention to content validity and pedagogical effectiveness. Whether this is an accurate interpretation should be examined in future research. Ideally, the builders of online learning resources will come to see usability as a necessary precondition for their sites just as test developers recognize that reliability is a necessary precondition before examining validity issues.

Site Builders Perceive Usability Evaluation as Important

The survey results clearly indicated that Web-based learning site builders considered usability evaluation important to the development of online learning resources. They would like to further improve the usability quality of their Web sites, and they think that, although usability is essential in online learning resources, it has been overlooked. They would also like to know more about usability evaluation, and think that

using an automated evaluation tool may be helpful in identifying usability issues in online learning resources (according to the Likert scale survey scores).

Although Web-based learning site builders think that usability evaluation is important, they have a generally low degree of knowledge about usability evaluation methods and seldom apply usability methods in developing online learning resources. The lack of organizational emphasis and support of usability may be the cause of this paradoxical phenomenon. Time and resource issues related to usability evaluation have surfaced in this study and in other usability-related literature (Flagg, 1990; Hughes & Burke, 2001; Ivory & Hearst, 2001; Kjeldskov, Skov, & Stage, 2004; Miller, 2005; Reeves & Hedberg, 2003).

It is also important to keep in mind that despite the fact that the Web is now 12 years old, it is still a relatively novel means of provided educational resources for most educators (Kearsley, 2005). In addition, the specific content of the DLESE resources is geosciences. Scientists and science educators who develop these resources are probably concerned first and foremost with the scientific accuracy of the content in the resources they develop and second with the pedagogical design of the resources. Usability may be a concern, but not as much as accuracy and pedagogy. Whether this interpretation is accurate should be the focus of future research.

Site Builders Perceive that Automated Usability Evaluation Has Limited Usefulness

The respondents agree that the usability evaluation results accurately reflect the usability quality of their Web-based learning sites. This finding is consistent with a previous study of the WebTango model, which predicted Web site usability quality with 81% and 94% accuracy (Ivory & Hearst, 2002). Although, it may be interesting to know

the usability quality ratings of a Web site, it would be much more useful if the ratings were more helpful for improving the sites. The respondents were undecided whether the usability evaluation results accurately reflect specific usability issues or whether they would be helpful for improving the usability quality of their Web sites. A previous study indicated that modifying Web sites using WebTango ratings did not significantly improve user performance or ratings of the Web sites (Ivory & Chevalier, 2002).

According to the open-ended comments, the Web-based learning site builders expressed some favorable perceptions of the automated usability evaluation results: 1) The results provide helpful information for improving usability of Web sites. 2) The results help pinpoint specific usability specific usability issues. 3) Using an automated tool is a helpful initial step(s) in developing a usability Web site. 4) The results and/or automated tools are more useful for novice Web builders.

The open-ended comments also indicated the following shortcomings of the usability evaluation results based on an automated usability evaluation: 1) The scope of the results is too limited and addresses only a subset of usability evaluation. In a related comment, the definition of usability, in automated usability evaluation tool, was said to be “misleading”. 2) The results summary provides too few details. The following comment from a respondent expresses the above two points well: “The current system lets me know if I’m in the ball park or not, but doesn’t tell me what I need to do to hit a home run”. 3) The result summary does not explain the specific issues and the associated usability principles of each issue. 4) The results summary does not provide information on how to improve usability of Web site. The following comment from a respondent sums up the above two points: “I cannot use these results to make specific definitive

changes to the website”. Basically, an automated usability evaluation tool is effective in rating Web site usability quality, but not effective in identifying issues or improving Web site usability quality.

In summary, this study’s finding, that Web-based learning sites have *average* or *below average* usability quality, is consistent with previous findings (Kirschner, Strijbos, Kreijns, & Beers, 2004; Norman & Draper, 1986; Storey, Phillips, Maczewski, & Wang, 2002). The builders of these Web-based learning sites possess a low degree of expertise in usability evaluation methods and seldom apply usability evaluation methods when developing online learning resources, perhaps because of limitations of time and money. Although the majority of the builders of these Web-based learning sites perceive usability evaluation as important in developing online learning resources, they do not as a rule do usability evaluation. The lack of usability evaluation is in keeping with the findings of the evaluation literature (Reeves & Hedberg, 2003).

An automated tool is effective in predicting Web site usability quality. However, the results of an automated usability evaluation tool have limited effectiveness in identifying specific Web site usability issues and improving Web site usability quality. The benefits and shortcomings of automated usability indicated in this study are basically consistent with the perceptions of other previously presented viewpoints regarding automated usability evaluation (Brajnik, 2004; Clapsaddle, 2004; Ivory, 2003; Laskowski, Landay, & Lister, 2002; Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004). In short, an automated usability evaluation tool is useful for limited aspects of usability evaluation within limited contexts.

Implications

Increasing Knowledge and Providing Support

Increasing Web-based learning site builders' knowledge about usability and providing support for them will likely increase the usability quality of online learning resources. According to the results of this study, the usability quality of Web-based learning sites has room for improvement. The site builders have a strong interest in usability, but they have little usability knowledge and seldom apply usability methods. Providing training and other resources for Web-based learning site builders will increase their knowledge, and providing support to practice usability evaluation will increase the usability quality of online learning resources.

Replacing the Word Usability with Words with more Specific Meanings

How a person defines the word usability has an impact on his or her perception and expectation of an *automated usability evaluation tool*. Using an alternate label for the word usability that has a more limited meaning may help establish a more realistic expectation of what a particular type of automated usability evaluation tool does. This concept is similar to a *spell checker* or a *grammar checker*. They are not called *writing checkers*. The tasks automated tools can do to enhance writing are relatively clear and limited (e.g., spelling and grammar). People do not expect an automated spell checker or grammar checker to analyze the quality of their *writing per se*.

Similar to a spell checker or a grammar checker, an automated usability evaluation does appear to offer some useful features. Having a narrower label for an automated evaluation tool may help Web site builders to better understand what an

automated usability evaluation tool has to offer and know when to (or not to) use an automated usability evaluation tool.

Offering Specific Feedback and Direction to Resolve the Issues

Offering specific feedback about a usability issue and specific direction on how to resolve the issue will allow an automated usability evaluation tool to be more *useful* for Web developers. An automated usability evaluation tool should identify the location of the issue. Then, the tool should present usability guidelines related to that particular issue.

Creating an automated usability evaluation tool that offers specific feedback and provides direction to resolve the issues is conceptually and technically feasible. This tool is similar to a spell checker that locates misspelled words and associates the findings to a dictionary. For example, one of the 157 items in the WebTango model is *Sans Serif Word Count* (see Table B-10). There is also a Font/Text Size entry in the Research-based Web Design & Usability Guidelines (Koyani, Bailey, & Nall, 2004) that describes the proper use of fonts on a Web page. An automated evaluation tool scans Web pages to accumulate a *Sans Serif Word Count* and then uses this count as a variable, among other ones, in analyzing the usability of a Web page. Therefore, the tool is capable of tracking the location of words with (or without) *Sans Serif* font and pages that have possible “font issues”. Such tools should be able to link to an entry that describes the font usage (e.g., Font/Text Size entry in the Research-based Web Design & Usability Guidelines (Koyani, Bailey, & Nall, 2004)). This back-end design would allow the necessary elements for developing an evaluation tool that 1) points out the specific locations of “font issues” and 2) directs the user of the tool to the solutions for resolving the issues.

Of course, the 157 measurement items of WebTango do not cover all aspects of Web usability evaluation. It is realistic to expect an automated tool to check a limited subset of Web usability evaluation, similar to the way a spell check validates a subset of writing quality--spelling. Basically, given the fact that the WebTango model has 157 items that can be used effectively to predict usability quality of a Web site, we should be able to combine WebTango (or other similar models) with a set of usability guidelines to create an automated “usability” evaluation tool that analyzes limited aspects of Web usability in a way that is more useful and usable to Web site builders.

The Implications of Expert vs. Novice Site Builders

One respondent made this comment: “Do usability surveys discriminate between the knowledgeable and non-knowledgeable users, and are separate usability metrics calculated?” Although this question is specifically related to the knowledge of users, it raises in my mind issues related to the knowledge and expertise of the site builders themselves. Of course, this dissertation was not designed to address the implication of expert versus novice site builders on usability evaluation, but the topic is worth discussing in this context.

Site builders’ usability knowledge level and Web site usability quality level may have an impact on interpreting the outcomes of a study and the direction of future research studies or “problem solving” approaches. Table 6-1 illustrates four possible combinations of site builders’ *usability knowledge level* and *Web site usability quality level*.

Table 6-1

The Four Combinations of Usability Knowledge Level and Web site Usability Quality Level

	Usability Knowledge Level	Web Site Usability Quality Level
Case 1	Low	Low
Case 2	Low	High
Case 3	High	Low
Case 4	High	High

The following are some implications of each of the cases in Table 6-1.

In case 1, the usability knowledge is *low* and the Web site usability quality is *low*.

Assuming that there is an interest in improving usability quality of the Web site, the question of why the knowledge is low would be addressed and the problem solved accordingly. The likely solution of this case is as follows: The organization would provide training and material to increase site builders' knowledge about usability. More importantly, the organization would provide the time and resources for site builders to apply their usability knowledge.

In case 2, the usability knowledge is *low* and the Web site usability quality is *high*. This is fortunate, considering that the quality is high despite the lack of knowledge. However, given that this result may be serendipitous, it still seems useful to recommend that the organization provide training related to usability to the developers.

In case 3, the usability knowledge is *high* and the Web site usability quality is *low*. A likely implication of this case is that the site builders are not given sufficient time and resources to incorporate usability evaluation into their development process. The theory would be verified and solutions provided accordingly.

In case 4, the usability knowledge is *high* and the Web site usability quality is *high*; this is the ideal situation.

There are other implications as well. For instance, research studies (Vanderdonckt et al., 2004) and some open-ended comments derived from this study suggest that by using an automated usability evaluation, novice Web designers may learn something useful, if rudimentary, about usability principles. Also, when designing an automated evaluation tool or heuristic evaluation guidelines, a possible consideration would be to provide more details to a novice than to an expert site builder. However, this must be subjected to further research.

In summary, there are reasons to believe that the level of expertise of a site builder does have implications for the way in which an automated usability evaluation method is applied.

Future Research Directions

Although automated usability evaluation tools have shown some promise, they are not at the point of making significant impact on improving Web site usability quality. Much more research and development is needed to refine automated usability evaluation tools, which may ultimately lead to a high degree of utilization. Clearly, these automated tools will not have a tangible impact on improving Web site usability quality without substantial investments in their development. This section describes the following suggested research areas and a framework for guiding research and development activities in these areas:

- The Current Status of Automated Usability Evaluation Tools
- The Features, Scope and Specificity of Explanation and Guidance

- The Impact of Usability and other Aspects of Web-based Learning Sites
- Multidisciplinary Development Research

The Current Status of Automated Usability Evaluation Tools

Investigating multiple current automated usability evaluation tools will help gain knowledge about the capability, limitations and future direction of automated evaluation tools. This dissertation and other studies confirm that automated usability evaluation tools are useful to a certain extent within certain contexts (Brajnik, 2004; Clapsaddle, 2004; Elizabeth & Tharam, 1997; Forsythe, 2003; Ivory, 2001, 2003; Laskowski, Landay, & Lister, 2002; Obendorf, Weinreich, & Hass, 2004; Vanderdonckt, Beirekdar, & Noirhomme-Fraiture, 2004). Automated usability evaluations are being studied in research environments and are being used in practical settings.

However, despite some advances, automated usability tools lack consistent purpose, functionalities, and operating procedures. Understandably, users have different expectations of what the tool does and how the tool should operate. In short, current automated usability tools are not useful enough for practitioners. Further investigations will enable the development of tools that are useful to practitioners. In order for automated usability approaches to have a tangible positive impact on Web site usability quality, we must develop automated usability evaluation tools that are useful. To gain a clear understanding of the capability, limitations, and future direction of automated evaluation, the following research questions should be pursued:

- What is the most current status of automated usability evaluation tools?
- What are the useful features of automated usability evaluation tools?
- What are the shortcomings of automated usability evaluation tools?

- What are the desired features of automated usability evaluation tools?
- What are the appropriate labels for automated usability evaluation tools?
- What should be the future direction of automated usability evaluation tools?

Given that the scope of the suggested study is at a relatively high level, it is suggested that a major corporation such as Google should invest in this type of research. Companies like Google would directly benefit from a more user-friendly Web. The results an intensive research and development initiative in this area might encourage other researchers to investigate the appropriate features, scope, and specificity of explanations and guidance provided by an automated usability evaluation tool.

The Features, Scope and Specificity of Explanation and Guidance

Investigating the appropriate features, scope, and specificity of explanations and guidance provided by an automated usability evaluation tool will improve the usefulness and usability of the tool. The goal of this research area should be to understand what a tool should do and how it should operate.

The survey respondents clearly wished that an automated usability evaluation tool would produce a more extensive and detailed explanation of the issues and provide detailed directions on how to resolve the issues. Having a clear understanding of the appropriate scope and level of detail of the evaluation results is essential for developing a useful and usable evaluation tool. This research area requires in-depth investigations of Web-site usability quality and practitioners' opinions. The following are the possible research questions of the suggested study:

- What usability measures should an automated evaluation tool offer?
- What operating features should an automated evaluation tool offer?

- What specificity of reported issues should an automated usability tool generate?
- What specificity of usability guidelines should an automated usability tool generate?
- How should an automated tool present the usability issues and guidelines?
- How should an automated usability evaluation tool adapt to various levels of users (novice vs. experts)?

The Impact of Usability and other Aspects of Web-based Learning Sites

Investigating the impact of usability on the effectiveness of a Web-based learning site (Kirschner, Strijbos, Kreijns, & Beers, 2004; Miller, 2005; Storey, Phillips, Maczewski, & Wang, 2002) is an essential aspect of the study of Web site usability quality. Web site effectiveness may involve various dimensions of measurements (e.g., pedagogy soundness, learning outcomes, or popularity measured by Google ranking).

The major research questions for this suggested research area are as follows:

- What is the correlation between usability and other effectiveness measurements of Web-based learning sites?
- How can usability standards best be established to guide the development of future Web-based learning sites?

Multidisciplinary Development Research

Development research (van den Akker, 1999) is an appropriate research framework for guiding further efforts in researching and developing automated usability evaluation tools so that a higher degree of utilization can be attained and so that these tools can actually have tangible benefits with respect to improving Web site usability quality. Development research has the following characteristics: 1) solving practical problems while maintaining high methodological rigor, 2) involving practitioners, and 3)

seeking and organizing generalized knowledge. Using development research as a framework will help to continually accumulate knowledge about Web usability applicable to the development of automated tool.

A multidisciplinary approach is also essential for developing automated usability evaluation tools. Although current automated usability evaluation tools are promising from a technical perspective, the usefulness and usability quality of these tools from users' perspective are questionable at best. For instance, WebTango was developed by a computer scientist affiliated with two of the best universities in the United States—The University of California at Berkeley and The University of Washington. WebTango analysis is based on 157 empirically validated measures. However, the steps for using WebTango are complicated, and interpreting the results is very difficult. Involving users, human learning and psychology experts, and user-interface design experts in researching and developing automated usability evaluation tools will likely enable the tools to be more usable and useful.

Summary

The goals of this dissertation were two-fold: 1) to gain insight into the current state of Web-based learning site usability based on an automated evaluation tool, and 2) to understand the perceptions of Web-based learning site builders about usability issues in general and the usefulness of usability analysis results generated by an automated tool in specific. The first goal was met by using WebTango to analyze Web-based learning sites. The usability quality of Web-based learning sites was rated by WebTango as average or below average. The second goal was met by conducting a survey study. The survey respondents knew relatively little about usability evaluation and rarely applied

these methods. In addition, they were generally uncertain regarding an automated evaluation tool. However, they also identified some benefits and issues of an automated evaluation tool that were helpful for future research and development of automated usability evaluation tools.

Automated usability evaluation clearly has a long way to go, and it is important to keep the expectations for automated tools at an appropriate level. With that in mind, I would like to end this dissertation with the voice of a survey respondent who summed up the role of an automated usability evaluation tool as follows:

While I think it is helpful to have a website reviewed by some type of automated usability evaluation, which I interpret as having a clear/static set of guidelines or standards to follow, I feel that, in regard to providing feedback, the automated system will fall short. I could see an automated system as being an initial step to a comprehensive usability evaluation. For example, do all the pages have titles, is the navigation consistent, etc. but because websites vary so drastically in content, visual presentation and so forth, and that they are designed for human interaction, I feel that there needs to be a human evaluation as well as an automated evaluation.

REFERENCES

- Academic Computing Center. (2005). Academic computing center. Retrieved April 8, 2005, from <http://www.coe.uga.edu/acc/>
- ACM. (2005). The ACM digital library. Retrieved June 30, 2005, from <http://portal.acm.org/dl.cfm>
- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). New York: Longman.
- Anderson, T., & Kanuka, H. (2003). *E-research: Methods, strategies, and issues* (1st ed.). Boston: Allyn and Bacon.
- Arms, W. Y. (2000). *Digital libraries*. Cambridge, MA: MIT Press.
- Babbie, E. R. (1973). *Survey research methods*. Belmont, CA: Wadsworth Publishing Company.
- Barab, S. A., MaKinster, J. G., Moore, J. A., & Cunningham, D. J. (2001). Designing and building an on-line community: The struggle to support sociability in the inquiry learning forum. *Educational Technology Research and Development*, 49(4), 71-96.
- Barstow, D., & Geary, E. (2002). *Blueprint for change: Report from the national conference on the revolution in earth and space science education*. Cambridge, MA: Technical Education Research Center.

- Bernard, M. (2000). Constructing user-centered websites: Design implications for content organization. *Usability News* 2(2).
- Bernard, M. (2001). Developing schemas for the location of common Web objects. *Usability News* 3(1).
- Bernard, M. (2002). User expectations for the location of common e-commerce Web objects. *Usability News* 4(1).
- Bernard, M., Baker, R., & Fernandez, M. (2002). Paging vs. Scrolling: Looking for the best way to present search results. *Usability News* 4(1).
- Bernard, M., & Hull, S. (2002). Where should you put the links? Comparing embedded and framed/non framed links. *Usability News* 4(1).
- Bernard, M., Liao, C., & Mills, M. (2001). Determining the best online font for older adults. *Usability News* 3(1).
- Bernard, M., Mills, M., Frank, T., & McKown, J. (2001). Which fonts do children prefer to read online? *Usability News* 3(1).
- Bias, R. G. (1994). The pluralistic usability walkthrough: Coordinated empathies. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 63-76). New York: Wiley.
- Bishop, A. P., & Bruce, B. C. (2002). *Usability research as participative inquiry*. Paper presented at the ACM+IEEE Joint Conference on Digital Libraries: Workshop on Usability of Digital Libraries, Portland, OR.
- Blackboard. (2005). Blackboard. Retrieved June 7, 2005, from <http://www.blackboard.com>

- Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). *Cognitive walkthrough for the web*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, Changing Ourselves, Minneapolis, MN.
- Bowlin, J. B. (2001). Linklint (Version 2.3.5).
- Brajnik, G. (2004). *Using automatic tools in accessibility and usability assurance*. Paper presented at the 8th ERCIM UI4ALL Workshop, Vienna.
- Bruning, R. H., Horn, C. A., & PytlikZillig, L. M. (2003). *Web-based learning: What do we know? Where do we go?* Greenwich, CT: Information Age.
- Carnevale, D. (2004a). MIT releases software for distance education. *The Chronicle of Higher Education*, 50.
- Carnevale, D. (2004b). More professors teach by using other colleges' online courses. *The Chronicle of Higher Education*, 51.
- Carroll, J. M. (2000). *Making use: Scenario-based design of human-computer interactions*. Cambridge, MA: MIT Press.
- Carroll, J. M. (2003). *HCI models, theories, and frameworks: Toward a multidisciplinary science*. San Francisco: Morgan Kaufmann.
- Carroll, J. M., & Rosson, M. B. (1992). Getting around the task-artifact cycle: How to make claims and design by scenario. *ACM Transactions on Information Systems*, 10(2), 181-212.
- Center for Information Technology Accommodation, U. S. G. S. A. (2002). Section 508. Retrieved June 1, 2005, from <http://section508.gov>

- Chalk, P. D. (2002). *Evaluating the use of a virtual learning environment for teaching aspects of HCI*. Paper presented at the Proceedings of the 7th Annual Conference on Innovation and Technology in Computer Science Education, Aarhus, Denmark.
- Clapsaddle, D. J. (2004). *Measuring usability: Categorically modeling successful websites using established metrics*. Unpublished Ph.D. Dissertation, Pace University, White Plains, NY
- Clark, R. C., & Mayer, R. E. (2003). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (1st ed.). San Francisco: Jossey-Bass/Pfeiffer.
- CMU. (2005). Distance education, ISRI. Retrieved June 8, 2005, from <http://www.distance.cmu.edu>
- Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. London: Sage Publications.
- deMarrais, K. B. (1998). *Inside stories: Qualitative research reflections*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Diaper, D. (1989). The discipline of HCI. *Interacting with Computers*, 1(1), 3-5.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley.
- Dix, A. (2003). *Human-computer interaction* (3rd ed.). New York: Pearson/Prentice-Hall.
- DLESE. (2005a). Digital library for earth system education. Retrieved June 10, 2005, from <http://www.dlese.org>

- DLESE. (2005b). Digital library for earth system education. Retrieved August 12, 2005, from <http://www.dlese.org>
- DLESE. (2005c). Educational resources overview. Retrieved June 10, 2005, from <http://www.dlese.org/resources/index.html>
- DLESE. (2005d). Educational resources: DLESE reviewed collection. Retrieved June 10, 2005, from <http://www.dlese.org/dds/histogram.do?group=resourceType>
- Doctor HTML. (2003). Doctor HTML (Version 6.1). Schenectady, NY: Imagiware, Inc.
- eCornell. (2005). eCornell. Retrieved June 8, 2005, from <http://www.ecornell.com>
- Elizabeth, C., & Tharam, S. D. (1997). *Automated usability testing*. Paper presented at the Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction.
- Evans, C., & Sabry, K. (2003). Evaluation of the interactivity of web-based learning systems: Principles and process. *Innovations in Education and Teaching International*, 40(1), 89-99.
- Finne, H., Levin, M., & Nilssen, T. (1995). Trailing research: A model for useful program evaluation. *Evaluation*, 1(1), 11-31.
- Flagg, B. N. (1990). *Formative evaluation for educational technologies*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Forsythe, A. M. (2003). *Development and evaluation of an automated approach to the measurement and design of icons*. Unpublished Ph.D. Dissertation, Queen's University of Belfast, Northern Ireland.
- Fowler, F. J. (1995). *Improving survey questions: Design and evaluation* (Vol. 38). Thousand Oaks, CA: Sage Publications.

- Fowler, F. J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Frank, E., Hall, M., Holmes, G., Martin, B., Pfahringer, B., Smith, T., et al. (2005). Weka (Version 3.4). Hamilton, New Zealand: Weka Machine Learning Project, The University of Waikato.
- Gagné, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design* (5th ed.). Belmont, CA: Thomson Learning, Inc.
- GALILEO. (2005). GALILEO: Georgia's virtual library. Retrieved June 30, 2005, from <http://www.galileo.usg.edu>
- Google. (2005a). Google. Retrieved June 30, 2005, from <http://www.google.com>
- Google. (2005b). Google scholar. Retrieved June 30, 2005, from <http://scholar.google.com>
- Granic, A., Glavinic, V., & Stankov, S. (2004). *Usability evaluation methodology for Web-based educational systems*. Paper presented at the 8th ERCIM Workshop "User Interfaces For All", Vienna.
- Hackos, J. T., & Redish, J. (1998). *User and task analysis for interface design*. New York: Wiley.
- Harmon, S. W., & Jones, M. G. (1999). The file levels of Web use in education: Factors to consider in planning online courses. *Educational Technology, XXXIX*(6), 28-32.
- Hines, T. (2000). An evaluation of two qualitative methods (focus group interviews and cognitive maps) for conducting research into entrepreneurial decision making. *Qualitative Market Research, 3*(1), 7-16.

- Hughes, M., & Burke, L. (2001). Usability testing of Web-based training. In B. H. Khan (Ed.), *Web-based training* (pp. 531-536). Englewood Cliffs, NJ: Educational Technology Publications.
- Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences* (3rd ed.). San Diego, CA: EdITS Publishers.
- ISO. (1998). ISO 9241-11:1998 ergonomic requirements for office work with visual display terminals (VDTs) -- part 11: Guidance on usability. Retrieved May 31, 2005, from <http://www.iso.org>
- ISO. (2003). ISO/ts 16071:2003 ergonomics of human-system interaction -- guidance on accessibility for human-computer interfaces. Retrieved May 31, 2005, from <http://www.iso.org>
- Ivory-Ndiaye, M. Y. (2004a). TangoViewer (Version 2.0). Seattle, WA: u b i t Research Program, University of Washington.
- Ivory-Ndiaye, M. Y. (2004b). WebTango analysis tool (Version 2.0). Seattle, WA: u b i t Research Program, University of Washington.
- Ivory, M. Y. (2001). *An empirical foundation for automated Web interface evaluation*. Unpublished Ph.D. Dissertation, UC Berkeley, Berkeley, CA.
- Ivory, M. Y. (2003). *Automated web site evaluation: Researcher's and practitioner's perspectives*. Dordrecht; Boston: Kluwer Academic Publishers.

- Ivory, M. Y., & Chevalier, A. (2002). *A study of automated Web site evaluation tools* (No. UW-CSE-02-10-01). Seattle, WA: Department of Computer Science and Engineering, University of Washington.
- Ivory, M. Y., & Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computer Survey*, 33(4), 470-516.
- Ivory, M. Y., & Hearst, M. A. (2002). *Statistical profiles of highly-rated Web sites*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves, Minneapolis, MN.
- Ivory, M. Y., Sinha, R. R., & Hearst, M. A. (2000). *Preliminary findings on using quantitative measures to compare favorably ranked and unranked information-centric Web pages*. Paper presented at the 6th Conference on Human Factors and the Web, Austin, TX.
- Ivory, M. Y., Sinha, R. R., & Hearst, M. A. (2001). *Empirically validated web page design metrics*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, WA.
- Johnson, J. (2003). *Web bloopers: 60 common web design mistakes, and how to avoid them*. San Francisco: Morgan Kaufmann.
- Jonassen, D. H. (Ed.). (2004). *Handbook of research on educational communications and technology* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Kahn, M. J., & Prail, A. (1994). Formal usability inspections. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 141-171). New York: Wiley.

- Kearsley, G. (Ed.). (2005). *Online learning: Personal reflections on the transformation of education*. Englewood Cliffs, NJ: Educational Technology Publications.
- Keith, S., Blandford, A., Fields, B., & Theng, Y. L. (2002). *An investigation into the application of claims analysis to evaluate usability of a digital library*. Paper presented at the ACM+IEEE Joint Conference on Digital Libraries: Workshop on Usability of Digital Libraries, Portland, OR.
- Kelly, B. (2002a). Bobby analysis of JISC 5/99 project entry points. Retrieved June 1, 2005, from <http://www.ukoln.ac.uk/qa-focus/surveys/web-10-2002/bobby>
- Kelly, B. (2002b). Web watch: An accessibility analysis of UK university entry points. *Ariadne* Retrieved June 1, 2005, from <http://www.ariadne.ac.uk/issue33/web-watch>
- Kelly, B. (2003a). CSS analysis of JISC 5/99 project entry points update. Retrieved June 1, 2005, from <http://www.ukoln.ac.uk/qa-focus/surveys/web-05-2003/css/>
- Kelly, B. (2003b). HTML analysis of JISC 5/99 project entry points update Retrieved June 1, 2005, from <http://www.ukoln.ac.uk/qa-focus/surveys/web-05-2003/html>
- Khan, B. H. (1997). *Web-based instruction*. Englewood Cliffs, NJ: Educational Technology Publications.
- Khan, B. H. (Ed.). (2001). *Web-based training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Kirakowski, J., Claridge, N., & Whitehand, R. (1998). *Human centered measures of success in Web site design*. Paper presented at the 4th Human Factors and the Web Conference, Basking Ridge, NJ.

- Kirschner, P., Strijbos, J.-W., Kreijns, K., & Beers, P. J. (2004). Designing electronic collaborative learning environments. *Educational Technology Research and Development*, 52(3), 71-96.
- Kjeldskov, J., Skov, M. B., & Stage, J. (2004). *Instant data analysis: Conducting usability evaluations in a day*. Paper presented at the Proceedings of the third Nordic conference on Human-computer interaction, Tampere, Finland.
- Kolodner, J. L. (1993). *Case-based reasoning*. San Mateo, CA: Morgan Kaufmann Publishers.
- Koyani, S. J., Bailey, R. W., & Nall, J. R. (2004). *Research-based Web design & usability guidelines: Computer Psychology*.
- Krug, S. (2000). *Don't make me think!: A common sense approach to Web usability*. Indianapolis, IN: New Riders.
- Laskowski, S. J., Landay, J. A., & Lister, M. (2002). *Automatic capture, representation, and analysis of user behavior*. Paper presented at the CHI '02 Extended Abstracts on Human Factors in Computing Systems, Minneapolis, MN.
- Lewis, C., Polson, P. G., Wharton, C., & Rieman, J. (1990). *Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People, Seattle, WA.
- Lohr, L., Javeri, M., Mahoney, C., Gall, J., Li, K., & Strongin, D. (2003). Using rapid application development to improve the usability of a preservice teacher technology course. *Educational Technology Research and Development*, 51(2), 41-55.

- Markham, J., & Hall, R. H. (2003). *E-commerce Web objects: Importance and expected placement* (No. LITE-2003-02). Rolla, MO: Laboratory for Information Technology Evaluation (LITE), University of Missouri-Rolla.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.
- Mazzarol, T., & Choo, S. (2003). A study of the factors influencing the operating location decisions of small firms. *Property Management* 21(2), 190-208
- Miller, M. J. (2005). Usability in e-learning. Retrieved June 8, 2005, from <http://www.learningcircuits.org/2005/jan2005/miller.htm>
- Molich, R., Thomsen, A. D., Karyukina, B., Schmidt, L., Ede, M., Oel, W. v., et al. (1999). *Comparative evaluation of usability tests*. Paper presented at the CHI '99 Extended Abstracts on Human Factors in Computing Systems, Pittsburgh, PA.
- NetLibrary. (2005). Netlibrary: The world's leading provider of ebooks. Retrieved June 30, 2005, from <http://www.netlibrary.com>
- Nielsen, J. (1993). *Usability engineering*. Boston: Morgan Kaufmann.
- Nielsen, J. (2000a). *Designing Web usability*. Indianapolis, IN: New Riders.
- Nielsen, J. (2000b). Why you only need to test with 5 users. *Alertbox: Current Issues in Web Usability* Retrieved June 23, 2005, from <http://www.useit.com/alertbox/20000319.html>
- Nielsen, J., & Landauer, T. K. (1993). *A mathematical model of the finding of usability problems*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Amsterdam.
- Nielsen, J., & Mack, R. L. (1994). *Usability inspection methods*. New York: Wiley.

- Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People, Seattle, WA.
- Nielsen, J., & Tahir, M. (2002). *Homepage usability: 50 websites deconstructed*. Indianapolis, IN: New Riders.
- Norman, D. A., & Draper, S. W. (1986). *User centered system design: New perspectives on human-computer interaction*. Hillsdale, NJ: L. Erlbaum Associates.
- Obendorf, H., Weinreich, H., & Hass, T. (2004). *Automatic support for web user studies with scone and tea*. Paper presented at the CHI '04 Extended Abstracts on Human Factors in Computing Systems, Vienna.
- Punch, K. (2003). *Survey research: The basics*. Thousand Oaks, CA: Sage Publications.
- Ratner, J. (2003). *Human factors and Web development* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Reeves, T. C., Apedoe, X., & Woo, Y. H. (2004). Evaluating digital libraries: A user-friendly guide. 2005, from <http://www.ucar.edu/>
- Reeves, T. C., & Carter, B. J. (2001). Usability testing and return-on-investment studies: Key evaluation strategies for Web-based training. In B. H. Khan (Ed.), *Web-based training* (pp. 531-536). Englewood Cliffs, NJ: Educational Technology Publications.
- Reeves, T. C., & Harmon, S. W. (1993). User interface rating tool for interactive multimedia. Retrieved June 5, 2005, from <http://it.coe.uga.edu/%7Etreeves/edit8350/UIRF.html>

- Reeves, T. C., & Hedberg, J. G. (2003). *Interactive learning systems evaluation*. Englewood Cliffs, NJ: Educational Technology Publications.
- Reigeluth, C., & Nelson, L. (1997). A new paradigm of ISD. In R. M. Branch & B. B. Minor (Eds.), *Educational media and technology yearbook* (Vol. 22, pp. 24–35). Englewood, CO: Libraries Unlimited.
- Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: Scenario-based development of human-computer interaction* (1st ed.). San Francisco: Academic Press.
- Rowley, D. E., & Rhoades, D. G. (1992). *The cognitive jogthrough: A fast-paced user interface evaluation procedure*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Monterey, CA.
- Schank, R. C. (1982). *Dynamic memory: A theory of reminding and learning in computers and people*. Cambridge: Cambridge University Press.
- Schank, R. C. (1998). Inside multi-media case based instruction. Mahwah, NJ: L. Erlbaum Associates.
- Schank, R. C., & Cleary, C. (1995). *Engines for education*. Hillsdale, NJ: L. Erlbaum Associates.
- Schank, R. C., Kass, A., & Riesbeck, C. K. (1994). *Inside case-based explanation*. Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Schonlau, M., Fricker, R. D., & Elliott, M. N. (2002). *Conducting research surveys via e-mail and the web*. Santa Monica, CA: Rand.
- Seidman, I. (1998). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (2nd ed.). New York: Teachers College Press.

- Setzer, J. C., & Lewis, L. (2005). *Distance education courses for public elementary and secondary school students: 2002-03*. Washington, DC National Center for Education Statistics, U.S. Department of Education.
- Sheard, J., & Markham, S. (2005). Web-based learning environments: Developing a framework for evaluation. *Assessment & Evaluation in Higher Education*, 30(4), 353-368.
- Sherman, P., & Quesenbery, W. (2005). Engineering the user experience: UX and the usability professionals' association. *Interactions*, 12(3), 38-40.
- Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd ed.). Reading, MA: Addison Wesley Longman.
- Spool, J., & Schroeder, W. (2001). *Testing web sites: Five users is nowhere near enough*. Paper presented at the CHI '01 Extended Abstracts on Human Factors in Computing Systems, Seattle, WA.
- SPSS. (2001). SPSS (Version 11.0). Chicago, IL: SPSS Inc.
- Sterne, J. (2002). *Web metrics: Proven methods for measuring Web site success*. New York: Wiley.
- Storey, M. A., Phillips, B., Maczewski, M., & Wang, M. (2002). Evaluating the usability of Web-based learning tools. *Educational Technology & Society*, 5(3), 91-100.
- Survey Research Center. (2005). Survey research center. Retrieved April 20, 2005, from <http://www.src.uga.edu>
- University of Georgia Libraries. (2005). Electronic journal locator. Retrieved June 30, 2005, from <http://www.libs.uga.edu/ejournals>
- UsableNet. (2004). Lift online. New York: UsableNet Inc.

- van den Akker, J. (1999). Principles and methods of development research. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 1-14). The Netherlands: Kluwer Academic Publishers.
- Vanderdonckt, J., Beirekdar, A., & Noirhomme-Fraiture, M. (2004). Automated evaluation of Web usability and accessibility by guideline review. *Lecture Notes in Computer Science*, 3140 17-30.
- Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors*, 34(4), 457-468.
- W3C. (2004). Web accessibility initiative. Retrieved June 1, 2005, from <http://www.w3.org/WAI>
- Waits, T., & Lewis, L. (2003). *Distance education at degree-granting postsecondary institutions: 2000-2001*. Washington, DC National Center for Education Statistics, U.S. Department of Education.
- Wang, F.-K., Moore, J. L., Wedman, J., & Shyu, C.-R. (2003). Developing a case-based reasoning knowledge repository to support a learning community—an example from the technology integration community. *Educational Technology Research and Development*, 51(3), 45-62.
- Wang, S.-K. (2003). *Development research with cognitive tools: An investigation of the effects of a web-based learning environment on student motivation and achievement in high school earth science*. Unpublished Ph.D., The University of Georgia, Athens, GA.
- Watchfire. (2004). WebXACT. Waltham, MA: Watchfire Corporation.

- Webby. (2005). Webby awards. Retrieved June 9, 2005, from <http://www.webbyawards.com>
- WebCT. (2005). WebCT. Retrieved June 5, 2005, from <http://www.webct.com>
- Weinman, L., & Karp, A. (2003). *Designing Web graphics.4* (4th ed.). Indianapolis, IN: New Riders.
- Wilkinson, D. (2000). *The researcher's toolkit: The complete guide to practitioner research*. London; New York: Routledge.
- Woolrych, A., & Cockton, G. (2001). *Why and when five test users aren't enough*. Paper presented at the IHM-HCI 2001 Conference, Lille, France.
- Wright, M., Marlino, M., & Sumner, T. (2002). Meta-design of a community digital library [Electronic Version]. *D-Lib Magazine*, 8(5). Retrieved August 11, 2005, from <http://www.dlib.org/dlib/may02/wright/05wright.html>
- Zhang, Z., Basili, V., & Shneiderman, B. (1999). Perspective-based usability inspection: An empirical validation of efficacy. *Empirical Software Engineering*, 4(1), 43.

APPENDICES

Appendix A: Automated Guideline Review Usability Evaluation Tools

Table A-1

Automated Guideline Review Usability Evaluation Tools (Ivory, 2003).

Tool Name	URL
508 Accessibility Suite	http://www.usablenet.com/frontend/508as_entry.jsp
AccessEnable	http://www.retroaccess.com/
AccVerify/AccRepair	http://www.hisoftware.com/access/
A-Prompt	http://aprompt.snow.utoronto.ca/
WatchFire Bobby	http://webxact.watchfire.com/
Dr. Watson	http://watson.addy.com/
LIFT	http://www.usablenet.com/
PageScreamer	http://www.crunchy.com/tools/PageScreamer.html
W3C CSS Validation Service	http://jigsaw.w3.org/css-validator/
W3C HTML Validation Service	http://validator.w3.org/
Weblint	http://search.cpan.org/~neilb/weblint-1.020/
WebSAT	http://zing.ncsl.nist.gov/WebTools/WebSAT/overview.html
WebTango	http://webtango.ischool.washington.edu/pages/tools.php

Appendix B: The WebTango Web Interfaces Usability Measurements

Table B-1

The 157 Quantitative Measures of WebTango for Evaluating Web Interfaces (Ivory, 2001).

Graphic Elements	Page Performance	Text Elements
Graphic Count	Table Count	Word Count
Graphic Link Count	HTML File Count	Page Title Word Count
Animated Graphic Count	HTML Bytes	Overall Page Title Word Count
Graphic Ad Count	Graphic File Count	Invisible Word Count
	Graphic Bytes	Meta Tag Word Count
Graphic Formatting	Script File Count	Body Word Count
Animated Graphic Ad Count	Script Bytes	Display Word Count
Redundant Graphic Count	Object File Count	Display Link Word Count
Minimum Graphic Height	Object Bytes	Link Word Count
Maximum Graphic Height	Object Count	Average Link Words
Average Graphic Height	Download Time	Graphic Word Count
Minimum Graphic Width	Bobby Approved	Ad Word Count
Maximum Graphic Width	Bobby Priority 1 Errors	Exclamation Point Count
Average Graphic Width	Bobby Priority 2 Errors	Spelling Error Count
Graphic Pixels	Bobby Priority 3 Errors	Good Word Count
	Bobby Browser Errors	Good Body Word Count
Link Elements	Weblint Errors	Good Display Word Count
Link Count	Visible Page Text Terms	Good Display Link Word Count
Text Link Count	Visible Unique Page Text Terms	Good Link Word Count
Link Graphic Count	Visible Page Text Hits	Average Good Link Words
Page Link Count	Visible Page Text Score	Good Graphic Word Count
Internal Link Count	All Page Text Terms	Good Page Title Word Count
Redundant Link Count	All Unique Page Text Terms	Overall Good Page Title Word Count
	All Page Text Hits	Good Meta Tag Word Count
Link Formatting	All Page Text Score	Reading Complexity
Non-underlined Text Links	Visible Link Text Terms	Overall Reading Complexity
Link Color Count	Visible Unique Link Text Terms	Fog Word Count
Standard Link Color Count	Visible Link Text Hits	Fog Big Word Count
	Visible Link Text Score	Overall Fog Big Word Count
Page Formatting	All Link Text Terms	Fog Sentence Count
Color Count	All Unique Link Text Terms	Overall Fog Sentence Count
Minimum Color Use	All Link Text Hits	
Browser-Safe Color Count	All Link Text Score	Text Formatting
Good Text Color Combination	Page Title Terms	Emphasized Body Word Count
Neutral Text Color Combinations	Unique Page Title Terms	Bolded Body Word Count
Bad Text Color Combinations	Page Title Hits	Capitalized Body Word Count
Good Panel Color Combinations	Page Title Score	Colored Body Word Count
Neutral Panel Color Combinations		Exclaimed Body Word Count
Bad Panel Color Combinations	Site Architecture	Italicized Body Word Count
Font Count	Text Element Variation	Underlined Word Count
Serif Font Count	Page Title Variation	Serif Word Count
Sans Serif Font Count	Link Element Variation	Sans Serif Word Count
Undetermined Font Style Count	Graphic Element Variation	Undetermined Font Style Word Count
Page Height	Text Formatting Variation	Font Style
Page Width	Link Formatting Variation	Minimum Font Size
Page Pixels	Graphic Formatting Variation	Maximum Font Size
Vertical Scrolls	Page Formatting Variation	Average Font Size
Horizontal Scrolls	Page Performance Variation	Body Color Count
Interactive Element Count	Overall Element Variation	Display Color Count
Search Element Count	Overall Formatting Variation	Text Positioning Count
External Stylesheet Use	Overall Variation	Text Column Count
Internal Stylesheet Use	Page Count	Text Cluster Count
Fixed Page Width Use	Maximum Page Depth	Link Text Cluster Count
Page Depth	Maximum Page Breadth	Border Cluster Count
Page Type	Median Page Breadth	Color Cluster Count
Self Containment		List Cluster Count
Spamming Use		Rule Cluster Count

Table B-2

WebTango Measures: Graphic Elements (Ivory, 2001).

Graphic Count	Total images
Graphic Link Count	Number of images that are links
Animated Graphic Count	Number of animated images
Graphic Ad Count	Number of images that possibly indicate ads
Animated Graphic Ad Count	Number of animated images that possibly indicate ads
Redundant Graphic Count	Number of images that point to the same image files as other images

Table B-3

WebTango Measures: Graphic Formatting (Ivory, 2001).

Minimum Graphic Height	Minimum image height (in pixels)
Maximum Graphic Height	Maximum image height (in pixels)
Average Graphic Height	Average image height (in pixels)
Minimum Graphic Width	Minimum image width (in pixels)
Maximum Graphic Width	Maximum image width (in pixels)
Average Graphic Width	Average image width (in pixels)
Graphic Pixels	Total page area covered by images (in pixels)

Table B-4

WebTango Measures: Link Elements (Ivory, 2001).

Link Count	Total number of links
Text Link Count	Number of text links
Link Graphic Count	Number of links that are images
Page Link Count	Number of links to other sections (i.e., anchors) within the page
Internal Link Count	Number of links that point to destination pages within the site
Redundant Link Count	Number of links that point to the same destination pages as other links on the page

Table B-5

WebTango Measures: Link Formatting (Ivory, 2001).

Non-underlined Text Links	Whether there are text links without visible underlines (0 – No; 1 – Yes)
Link Color Count	Number of colors used for text links
Standard Link Color Count	Number of default browser colors used for text links

Table B-6

WebTango Measures: Page Formatting (Ivory, 2001).

Color Count	Number of colors used
Minimum Color Use	Minimum number of times a color is used
Browser-Safe Color Count	Number of browser-safe colors used
Good Text Color Combination	Number of good text or thin line color combinations
Neutral Text Color Combinations	Number of neutral text or thin line color combinations
Bad Text Color Combinations	Number of bad text or thin line color combinations
Good Panel Color Combinations	Number of good thick line or panel color combinations
Neutral Panel Color Combinations	Number of neutral thick line or panel color combinations
Bad Panel Color Combinations	Number of bad thick line or panel color combinations
Font Count	Number of fonts used
Serif Font Count	Number of serif font faces used
Sans Serif Font Count	Number of sans serif font faces used
Undetermined Font Style Count	Number of undetermined font faces used
Page Height	Height of page in pixels (600 pixel screen height)
Page Width	Width of page in pixels (800 pixel screen width)
Page Pixels	Total screen area required to render the page
Vertical Scrolls	Number of vertical scrolls required to view the entire page
Horizontal Scrolls	Number of horizontal scrolls required to view the entire page
Interactive Element Count	Number of text fields, radio boxes, buttons and other form objects
Search Element Count	Number of forms for performing a site search
External Stylesheet Use	Whether an external stylesheet file is used to format the page (0 – No; 1 – Yes)
Internal Stylesheet Use	Whether an internal stylesheet file is used within the head tag to format the page (0 – No; 1 – Yes)
Fixed Page Width Use	Whether tables are used to create a specific page width (0 – No; 1 – Yes)
Page Depth	Level of the page within the site
Page Type	Functional type (e.g., 0 -- home, 1 -- index, 2 -- content, 3 -- form, or 4 -- other)
Self Containment	The degree to which all page elements are rendered via the HTML and image files vs. scripts, external stylesheets, objects, etc. (0=low, 1=medium, and 2=high)
Spamming Use	Whether the page uses invisible text or long page titles (0 – No; 1 – Yes)

Table B-7

WebTango Measures: Page Performance (Ivory, 2001).

Table Count	Number of HTML tables used to render the page
HTML File Count	Number of HTML files, including stylesheet files
HTML Bytes	Total bytes for HTML tags, text, and stylesheet tags
Graphic File Count	Number of image files
Graphic Bytes	Total bytes for image files
Script File Count	Number of script files
Script Bytes	Total bytes for scripts (embedded in script tags and in script files)
Object File Count	Number of object files (e.g., for applets, layers, sound, etc.)
Object Bytes	Total bytes for object tags and object files (e.g., for applets, layers, sound, etc.)
Object Count	Number of scripts, applets, objects, etc.
Download Time	Time for a page to fully load over a 56.6K modem (41.2K connection speed)
Bobby Approved	Where the page was approved by Bobby as being accessible to people with disabilities (0 – No; 1 – Yes)
Bobby Priority 1 Errors	Number of Bobby priority 1 errors reported
Bobby Priority 2 Errors	Number of Bobby priority 2 errors reported
Bobby Priority 3 Errors	Number of Bobby priority 3 errors reported
Bobby Browser Errors	Number of Bobby browser compatibility errors reported
Webint Errors	Number of HTML syntax errors reported by Webint
Visible Page Text Terms	Maximum good visible words on source and destination pages
Visible Unique Page Text Terms	Maximum good visible, unique words on source and destination pages
Visible Page Text Hits	Common visible words on source and destination pages
Visible Page Text Score	Score for common visible words on source and destination pages
All Page Text Terms	Maximum good visible and invisible words on source and destination pages
All Unique Page Text Terms	Maximum good visible and invisible, unique words on source and destination pages
All Page Text Hits	Common visible and invisible words on source and destination pages
All Page Text Score	Score for common visible and invisible words on source and destination pages
Visible Link Text Terms	Maximum good visible words in the link text and destination page
Visible Unique Link Text Terms	Maximum good visible, unique words in the link text and destination page
Visible Link Text Hits	Common visible words in the link text and destination page
Visible Link Text Score	Score for common visible words in the link text and destination page
All Link Text Terms	Maximum good visible and invisible words in the link text and destination page
All Unique Link Text Terms	Maximum good visible and invisible, unique words in the link text and destination page
All Link Text Hits	Common visible and invisible words in the link text and destination page
All Link Text Score	Score for common visible and invisible words in the link text and destination page
Page Title Terms	Maximum good page title on source and destination pages
Unique Page Title Terms	Maximum good, unique page title words on source and destination pages
Page Title Hits	Common page title words on source and destination pages
Page Title Score	Score for common page title words on source and destination pages

Table B-8

WebTango Measures: Site Architecture (Ivory, 2001).

Text Element Variation	Variance in text elements across pages
Page Title Variation	Variance in page titles across pages
Link Element Variation	Variance in link elements across pages
Graphic Element Variation	Variance in graphic elements across pages
Text Formatting Variation	Variance in text formatting across pages
Link Formatting Variation	Variance in link formatting across pages
Graphic Formatting Variation	Variance in graphic formatting across pages
Page Formatting Variation	Variance in page formatting across pages
Page Performance Variation	Variance in page performance across pages
Overall Element Variation	Variance in all elements across pages
Overall Formatting Variation	Variance in all formatting across pages
Overall Variation	Variance in elements, formatting, and performance across pages
Page Count	Number of crawled pages
Maximum Page Depth	Maximum crawl depth
Maximum Page Breadth	Maximum pages crawled at a level
Median Page Breadth	Median pages crawled at a level

Table B-9

WebTango Measures: Text Elements (Ivory, 2001).

Word Count	Total visible words; also used for Reading Complexity over all text
Page Title Word Count	Number of words in the page's title (max of 64 chars)
Overall Page Title Word Count	Total number of words in the page's title (no character max)
Invisible Word Count	Number of invisible words
Meta Tag Word Count	Number of words in meta tags (key word & description)
Body Word Count	Words that are body text
Display Word Count	Words that are display text (i.e., headings that are not links)
Display Link Word Count	Words that are both display and link text (i.e., link headings)
Link Word Count	Words that are link text
Average Link Words	Average number of words in link text
Graphic Word Count	Number of words from image alt text tags
Ad Word Count	Number of words possibly indicating ads (e.g., advertisement or sponsor)
Exclamation Point Count	Number of exclamation points
Spelling Error Count	Number of spelling errors
Good Word Count	Total good visible words
Good Body Word Count	Good body text words (i.e., not stop words)
Good Display Word Count	Good display text words (i.e., not stop words)
Good Display Link Word Count	Good combined display and link text words (i.e., not stop words)
Good Link Word Count	Good link text words (i.e., not stop words or 'click')
Average Good Link Words	Average number of good words in link text (i.e., not stop words or 'click')
Good Graphic Word Count	Number of good words from image alt text tags
Good Page Title Word Count	Number of good page title words (max of 64 chars)
Overall Good Page Title Word Count	Total number of good page title words (no character max)
Good Meta Tag Word Count	Number of good meta tag words (i.e., not stop words)
Reading Complexity	Gunning Fog Index (ratios of words, sentences and words with 3+ syllables); computed over prose
Overall Reading Complexity	Gunning Fog Index (ratios of words, sentences and words with 3+ syllables); computed over all text
Fog Word Count	Number of prose words (for Reading Complexity)
Fog Big Word Count	Number of big prose words (for Reading Complexity)
Overall Fog Big Word Count	Number of big words (for Reading Complexity over all text)
Fog Sentence Count	Number of sentences for prose words (for Reading Complexity measure over prose text)
Overall Fog Sentence Count	Number of sentences (for Reading Complexity measure over all text)

Table B-10

WebTango Measures: Text Formatting (Ivory, 2001).

Emphasized Body Word Count	Body text words that are emphasized (e.g., bolded, capitalized, near !'s)
Bolded Body Word Count	Body text words that are bolded
Capitalized Body Word Count	Body text words that are capitalized
Colored Body Word Count	Body text words that are a color other than the default text color
Exclaimed Body Word Count	Body text words that are near exclamation points
Italicized Body Word Count	Body text words that are italicized
Underlined Word Count	Number of words that are underlined but are not links
Serif Word Count	Number of words formatted with serif font faces
Sans Serif Word Count	Number of words formatted with sans serif font faces
Undetermined Font Style Word Count	Number of words formatted with undetermined font faces
Font Style	Whether text is predominately sans serif, serif, or undetermined font styles
Minimum Font Size	Smallest font size (in points) used for text
Maximum Font Size	Largest font size (in points) used for text
Average Font Size	Predominate font size (in points) used for text
Body Color Count	Number of colors used for body text
Display Color Count	Number of colors used for display text
Text Positioning Count	Number of text areas that change position from flush left
Text Column Count	Number of x positions (i.e., columns) where text starts
Text Cluster Count	Number of text areas that are highlighted in some manner (with color, bordered regions, rules, or lists)
Link Text Cluster Count	Number of link text areas that are highlighted in some manner (with color, bordered regions, rules, or lists)
Border Cluster Count	Number of text and link text areas that are highlighted with bordered regions
Color Cluster Count	Number of text and link text areas that are highlighted with colored regions
List Cluster Count	Number of text and link text areas that are highlighted with lists
Rule Cluster Count	Number of text and link text areas that are highlighted with horizontal or vertical rules

Appendix C: WebTango Tools

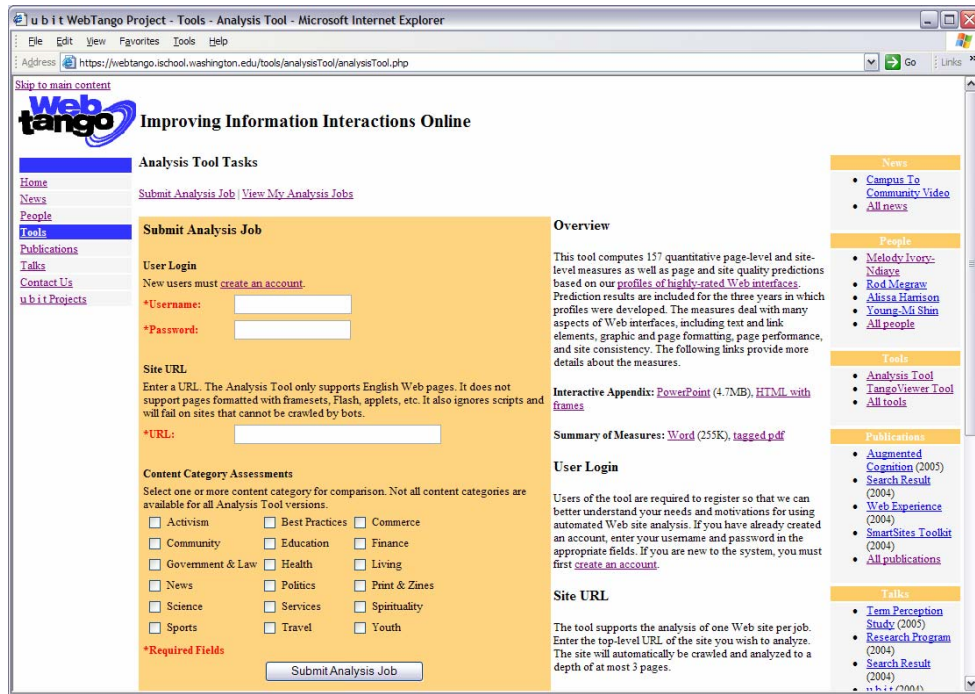


Figure C.1. WebTangle Analysis tool screen sample (Ivory-Ndiaye, 2004b).

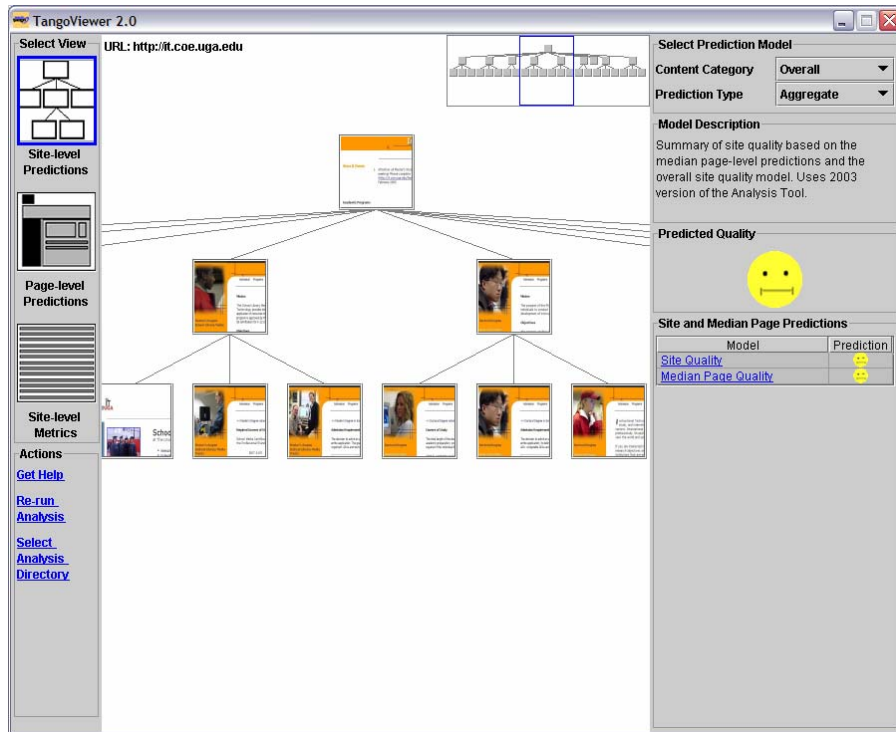


Figure C.2. TangoViewer tool screen sample (Ivory-Ndiaye, 2004a).

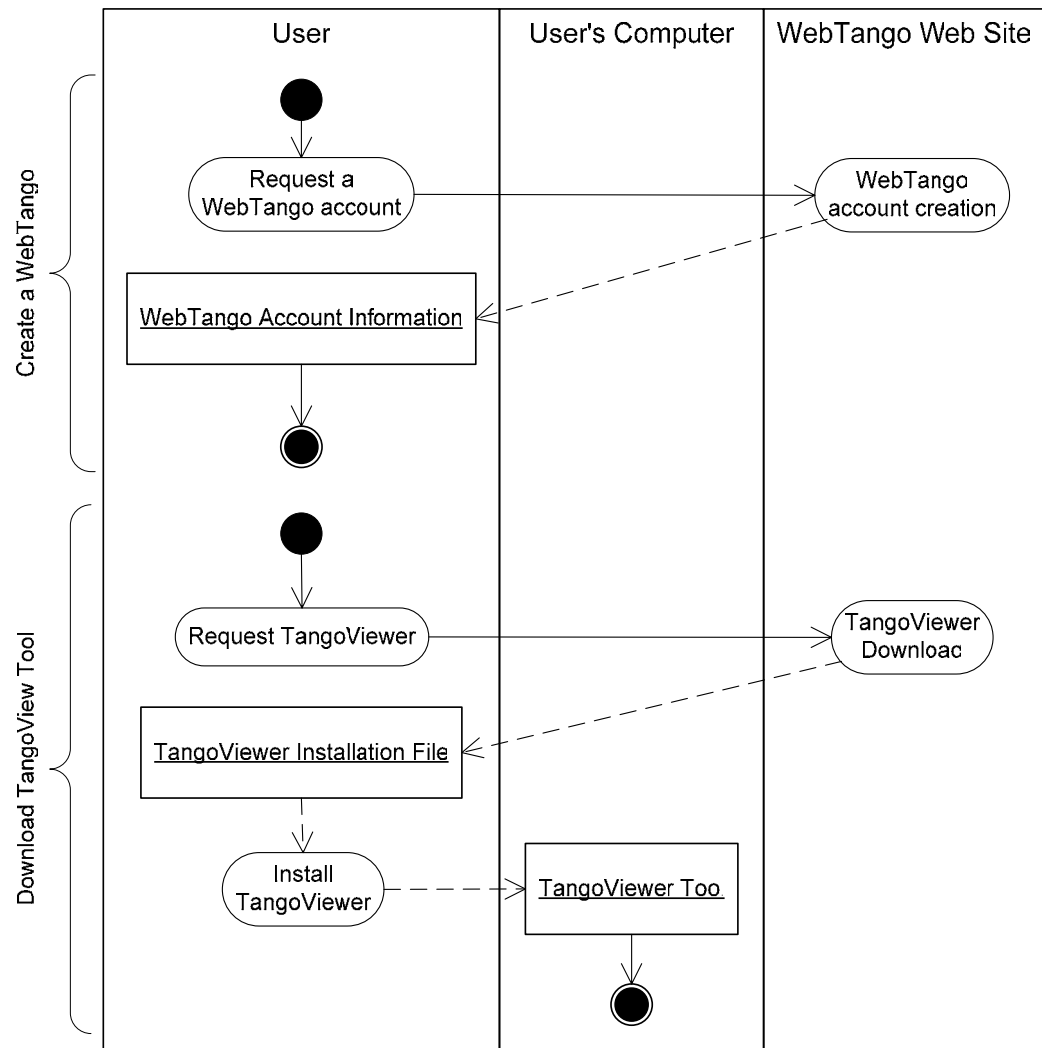


Figure C.3. WebTango initial setup process.

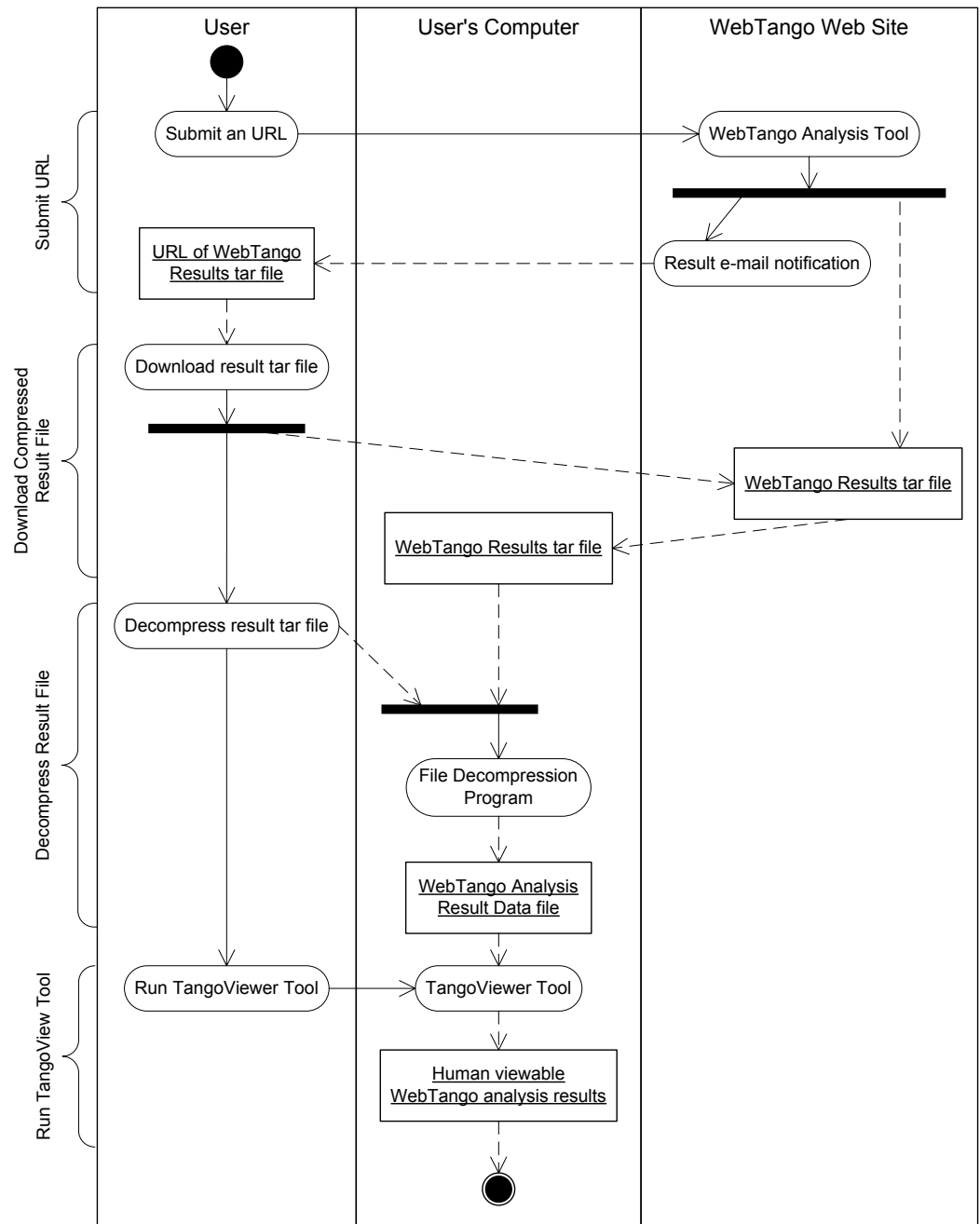


Figure C.4. WebTango analysis process.

Appendix D: Sample Excluded Sites

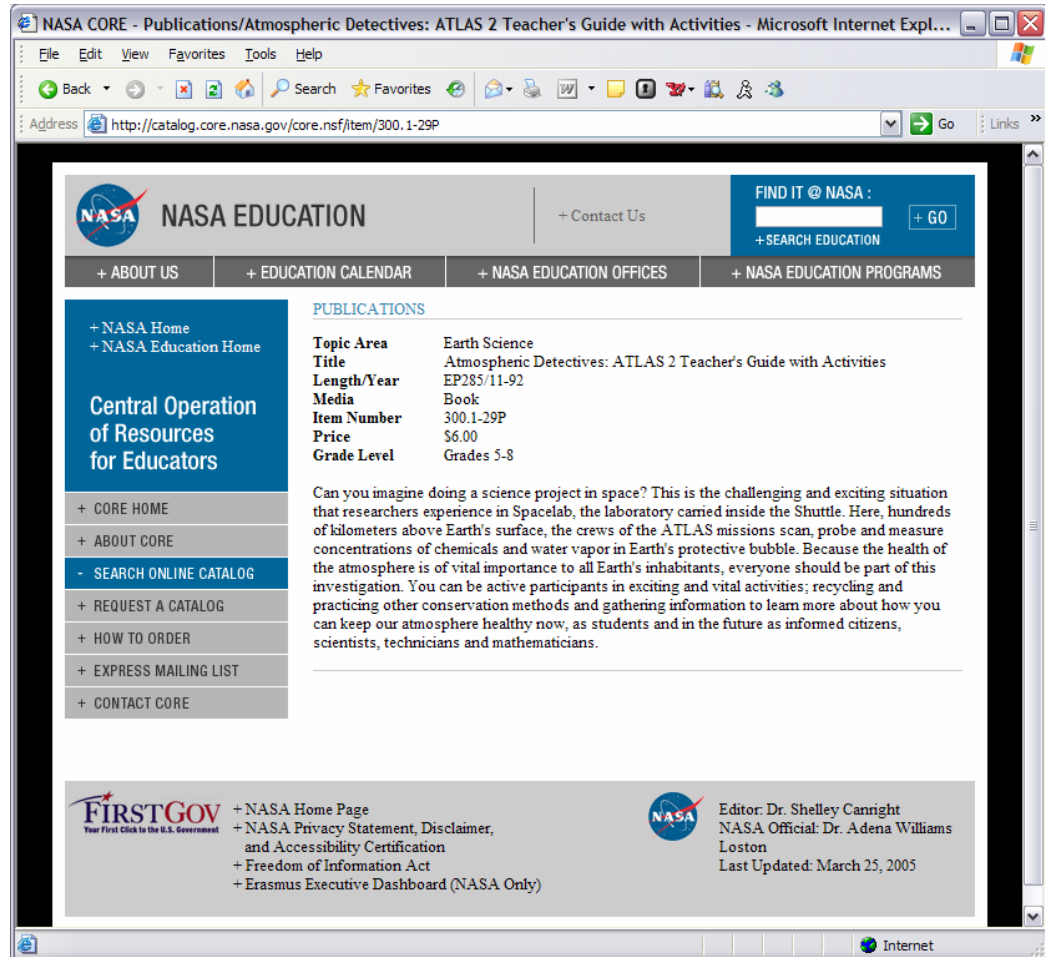


Figure D.1. A site that is a description of an actual learning material.

Daily Lesson Plan - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Print Mail

Address <http://www.nytimes.com/learning/teachers/lessons/20000127thursday.html> Go Links

The New York Times
LearningNetwork

Teacher
CONNECTIONS

[Back to Main](#)

[Daily Lesson Plan](#)
[Lesson Plan](#)
[Archive](#)
[News Snapshot](#)
[Issues in Depth](#)
[On This Day in History](#)
[Crossword Puzzle](#)
[Campus Weblines](#)
[Education News](#)
[Newspaper in Education \(NIE\)](#)
[Teacher Resources](#)
[Classroom](#)
[Subscriptions](#)

Student
CONNECTIONS

[News Summaries](#)
[Daily News Quiz](#)
[Word of the Day](#)
[Test Prep Question of the Day](#)
[Web Explorer](#)
[Science Q & A](#)

DAILY LessonPlan

Developed in Partnership with the
Bank Street College of Education in NYC

SEND THIS PAGE TO A FRIEND GET OUR LESSONS BY E-MAIL PRINTER-FRIENDLY VERSION

Thursday, January 27, 2000

Warning Signs for Dangerous Times

Exploring the Use of Storm-Tracking Technologies

Author(s)
[Catherine Wilson Bazant, The New York Times Learning Network](#)

Grades: 6-8, 9-12
Subjects: Geography, Language Arts, Technology
[Interdisciplinary Connections](#)


Overview of Lesson Plan: In this lesson, students explore the use of storm-tracking technologies, research and present how they forecast natural disasters, and assess the importance of these technologies.
Review the [Academic Content Standards](#) related to this lesson.

Suggested Time Allowance: 45 minutes-1 hour

Objectives:
Students will:


1. Imagine being in the midst of a natural disaster, and analyze what resources are needed to prepare for it.
2. Examine how and why storm-tracking technology is important by reading and discussing "Classic Northeast, Worst in 4 years, Leaves Six Dead."
3. Research various storm tracking technologies to determine how they forecast storms.

Related Article
[Classic Northeast, Worst in 4 Years, Leaves Six Dead](#)
By DAVID BARSTOW




[\(Go to Article.\)](#)

The affordable, expressive
HP Color LaserJet 3500n
and 3550n printers make
learning come to life.



» [Learn more](#)
» [Free tie-dye t-shirt offer](#)

Now, get **\$100** in
instant savings.
Offer ends October 31, 2005.



Internet

Figure D.2. Learning material that is part of a large portal.

Appendix E: Survey Instruments

The First Instrument: Addressing Research Questions #2.1, #2.2, and #2.3

Dear John:

You are invited to participate in a Web-based survey regarding the awareness of the developers of online learning resources with respect to usability evaluation methods and their practices in evaluating the usability of these learning resources.

This survey is part of my dissertation study, titled "Investigating Usability of Web-Based Learning Sites Using an Empirical Automated Usability Evaluation Approach and the Perceptions of Site Builders". My dissertation is under supervision of Dr. Thomas C. Reeves of the University of Georgia (<http://it.coe.uga.edu/~treeves>).

This invitation is being sent to people who identified themselves as the contact person (i.e., designer, owner, etc.) of Web sites accessed via the DLESE Reviewed Collection (DRC) of the Digital Library for Earth System Education (DLESE). Your identity will not be associated with the findings nor will it be released. This study has been approved by the Human Subjects Office of the University of Georgia (<http://www.ovpr.uga.edu/hso>).

Because your time is valuable, I limited this Web-based survey to only 12 items requesting your degree of agreement or disagreement. We would greatly appreciate it if you would complete the surveys by THURSDAY, NOVEMBER 10 at 3 P.M. This will allow us to meet the required semester deadlines. (Of course, we are still interested in your responses should you miss the deadline, so please complete the survey and forward any responses after the date above as well.)

After you have completed this survey, I will send you a summary of the usability evaluation results of <http://www.sample.com/> based upon an empirical-based automated usability evaluation tool, as well as a brief (4 items) follow-up Web-based survey. This second survey seeks your perceptions of how useful you view the results of an automated usability evaluation tool in identifying Web site usability issues and improving Web site usability quality.

I hope you will decide to contribute a small amount of your time to this study that ultimately could lead to the enhancement of Web-based learning. As an extra incentive, you will have a chance to win a \$100 gift certificate from www.amazon.com by completing both surveys. University of Georgia employees will be excluded from this extra incentive.

I would be pleased to answer any additional questions you may have about this study. Thank you for your contribution to this study.

Please go to <http://www.web4success.org/survey/index.php?sid=1&token=9324869412> to start the survey.

Sincerely,

Saun Shewanown
Doctoral Candidate
Instructional Technology, The University of Georgia
saun@uga.edu, 404-437-6356

Figure E.1. The invitation for the first survey instrument.

A Survey Study (Part One): Understanding of the Awareness, Perceptions and Practice in Evaluating Usability of Online Learning Resources

Welcome,

You are invited to participate in a research study titled "Investigating Usability of Web-Based Learning Sites Using an Empirical Automated Usability Evaluation Approach and the Perceptions of Site Builders" conducted by Saun Shewanown, Department of Educational Psychology and Instructional Technology, University of Georgia, 404-437-6356 under the direction of Dr. Thomas C. Reeves, Department of Educational Psychology and Instructional Technology, University of Georgia, 603D Aderhold Hall, Athens, GA 30602-7144.

This research study involves two brief surveys. The purpose of this part-one survey is to gain a better understanding of the awareness, perceptions and practice in evaluating the usability of online learning resources such as the one with which you are associated. After this Part One (12-question) survey is completed, you will be invited to participate in the Part Two survey that has 4 questions.

In the following screens, you will be asked to respond to a brief survey questionnaire that includes questions related to [usability testing](#), [usability inspection](#) (or heuristic evaluation) and [automated usability evaluation](#). There are no demographic related questions.

Completion of the survey is expected to take a maximum of 10 minutes. Please note that Internet communications are insecure and there is a limit to the confidentiality that can be guaranteed due to the technology itself. However, once I receive the completed surveys, I will store the data in a locked cabinet in my office and will destroy them and any names and contact information that I have by June, 2006. Any information that is obtained in connection with this study and that can be identified with you will remain confidential except as required by law. If you are not comfortable with the level of confidentiality provided by the Internet, please feel free to print out a copy of the survey, fill it out by hand, and mail it to me at the address given below, with no return address on the envelope.

Your participation in this study is completely voluntary. You may withdraw at any time without penalty, or skip any questions you feel uncomfortable answering. Closing the survey window will erase your answers without submitting them. Additionally, you will be given a choice of submitting or discarding your responses at the end of the survey. If you have any questions, please do not hesitate to ask now or at a later date. You may contact Saun Shewanown at 404-437-6356 or saun@uga.edu.

Thank you for the invaluable help that you are providing by participating in this research study.

Sincerely,

Saun Shewanown
EPIT Department, University of Georgia
604 Aderhold Hall, Athens, GA 30602-7144
saun@uga.edu, 404-437-6356

By completing the survey you are agreeing to participate in the research.
Please begin the survey **below**.
Click on the **Submit** button at the bottom of this page when you are done.

Figure E.2. The informed consent form for the first survey instrument.

Please check your **level of expertise** with the following [usability examination methods](#) using a scale where 1 = "Not familiar with it" to 5 = "expert in its application."

#1. Usability Testing: User(s) perform predetermined tasks while being observed and recorded.

#2. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.

#3. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.

Please check your **use** of the following [usability examination methods](#) using a scale where 1 = "Never" to 5 = "always."

#4. Usability Testing: User(s) perform predetermined tasks while being observed and recorded.

#5. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.

#6. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.

Figure E.3. Survey questions #1 through #6 addressing research questions #2.1 and #2.2.

Please check your agreement with each of the following statements using the scale below.
1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

#7. I am interested in further improvement of the usability of my Web site.

#8. Usability is an essential feature of online learning resources.

#9. Usability is usually overlooked in the design of online learning resources.

#10. Personally, I would like to know more about usability in the context of online learning resources.

#11. It would be very helpful to automate the process of finding usability problems in online learning resources.

#12. Personally, I am skeptical about the value of automating the process of finding usability problems in online learning resources.

Figure E.4. Survey questions #7 through #12 addressing research question #2.3.

Usability Evaluation Methods

Usability is defined 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” (Rossett, 2002).

Usability testing of a Web site is a process wherein the intended users of the site perform a set of predetermined tasks on the site while the users' behaviors are being observed by evaluators and recorded. A usability test may be conducted in a formal usability lab equipped with computers and audio and video recording devices; such a lab typically has two rooms, one for the user, the other for the evaluator(s) (Hughes & Burke, 2001). Usability tests can also be conducted in real-life settings, e.g., a classroom, using portable devices (Reeves, 1999). Usability testing is also known as formal usability testing, user testing, or portable usability testing).

Usability inspection (or Heuristic evaluation) is a process whereby evaluators examine the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria. Evaluators independently explore a Web site and judge the site, pages and elements based on a set of criteria. Heuristic evaluation criteria tend to be at a high-level (e.g., user control and freedom, or error prevention). When the evaluation is completed, a list of possible usability issues of the site is created based on the results of all evaluators. Usability inspection is also known as heuristic evaluation, cognitive walkthrough, perspective-based inspection, feature inspection, or guideline review.

An automated Web site usability evaluation tool is a software program that analyzes Web site user interfaces to determine the usability quality. Some tools are fully automated, others are partially automated. The main motivation for automating Web site usability evaluations is to reduce the amount of usability testing performed manually. Automated evaluation serves a similar purpose as heuristic evaluation and has many potential benefits.

Figure E.5. The usability evaluation method descriptions accompanying survey questions #7 through #12.

The Second Instrument: Addressing Research Question #3

[Home](#)

Usability Evaluation Results Summary

<http://www.sample.com>

Dear John:

Thank you very much for participating in the first part of my survey study. Below are the summarized results of using an automated Web usability evaluation tool to analyze <http://www.sample.com>. We understand that your time is valuable. Therefore, only **selected results** are presented here. This summary is **not intended to be a full usability evaluation**.

The automated Web usability evaluation tool results are summarized according four categories (Text, Graphic, Link, and Page). Each category has two sub-categories. There are three potential ratings for each sub-category: Excellent, Good, and Room for Improvement. Please note that the tool does not return results for every sub-category depending upon the nature of the specific site.

Evaluation Category	Sub-category	Rating
Text	Text Element	Not rated
	Text Formatting	Room for Improvement
Graphic	Graphic Element	Good
	Graphic Formatting	Excellent
Link	Link Element	Not rated
	Link Formatting	Room for Improvement
Page	Page Title	Good
	Page Formatting	Excellent

You can also display all the results on one page by clicking [Show All Categories](#).

After you have looked at each of the evaluation results and the accompanying usability guidelines, **please complete the accompanying Web-based survey** which constitutes the second and final part of my survey study. The **goal** of this survey is to learn about whether you think the **kind of results** generated by an automated Web usability evaluation tool is helpful to you for designing Web sites.

- Please go to <http://www.web4success.org/survey/index.php?sid=2&token=6427763175> to start the survey.

If you have any questions, please contact Saun Shewanown at saun@uga.edu or 404-437-6356.

Figure E.6. A sample WebTango analysis results.

[Home](#)

Usability Evaluation Results Summary

<http://www.sample.com>

Text

[How to Read the Evaluation Results](#)

Text_Formatting	Evaluation Results	Text-Related Usability Guidelines
Text Formatting refers to "how body text is emphasized; whether some underlined text is not in text links; how text areas are highlighted; font styles and size; number of text colors; number of times text is repositioned" (Ivory, 2003, p. 71).	Room for Improvement The results suggest that Text_Formatting aspects of http://www.sample.com may have some possible shortcomings. However, Web designers often have good contextual reasons to overrule generalized usability guidelines. Conducting a heuristic evolution based on Text-related usability guidelines may help determine whether the perceived Text_Formatting shortcomings should be addressed.	<ul style="list-style-type: none">• Text-Related Usability Guidelines at usability.gov• Reading-Related Usability Guidelines at usability.gov

After you have looked at each of the evaluation results and the accompanying usability guidelines, **please complete the accompanying Web-based survey** which constitutes the second and final part of my survey study. The **goal** of this survey is to learn about whether you think the **kind of results** generated by an automated Web usability evaluation tool is helpful to you for designing Web sites.

- Please go to <http://www.web4success.org/survey/index.php?sid=2&token=6427763175> to start the survey.

If you have any questions, please contact Saun Shewanown at saun@uga.edu or 404-437-6356.

Figure E.7. Sample text results summary.

Usability Evaluation Results Summary

<http://www.sample.com>

Graphic

[How to Read the Evaluation Results](#)

Graphic_Element	Evaluation Results	Graphic-Related Usability Guidelines
Graphic Element refers to the "number and type of images" (Ivory, 2003, p. 71) used on a Web site.	Good The results suggest that Graphic_Element aspects of http://www.sample.com are good and comparable with other similar sites. You may consider consulting the Graphic related usability guidelines for a few Graphic_Element guidelines that may have been overlooked.	<ul style="list-style-type: none">• Graphic-Related Usability Guidelines at usability.gov
Graphic Formatting refers to the "minimum, maximum, and average image width and height; page area covered by images" (Ivory, 2003, p. 71) used on a Web site.	Excellent The results suggest that Graphic_Formatting aspects of http://www.sample.com are significantly better than other similar sites. Feel free to consult the Graphic related usability guidelines for helpful information about Graphic_Formatting usability.	

After you have looked at each of the evaluation results and the accompanying usability guidelines, **please complete the accompanying Web-based survey** which constitutes the second and final part of my survey study. The **goal** of this survey is to learn about whether you think the **kind of results** generated by an automated Web usability evaluation tool is helpful to you for designing Web sites.

- Please go to <http://www.web4success.org/survey/index.php?sid=2&token=6427763175> to start the survey.

If you have any questions, please contact Saun Shewanown at saun@uga.edu or 404-437-6356.

Figure E.8. Sample graphic results summary.

[Home](#)

Usability Evaluation Results Summary

<http://www.sample.com>

Link

[How to Read the Evaluation Results](#)

Link_Formatting	Evaluation Results	Link-Related Usability Guidelines
Link Element refers to the "colors used for links and whether there are text links that are not underlined or colored" (Ivory, 2003, p. 71) used on a Web site.	Room for Improvement The results suggest that Link_Formatting aspects of http://www.sample.com may have some possible shortcomings. However, Web designers often have good contextual reasons to overrule generalized usability guidelines. Conducting a heuristic evolution based on Link-related usability guidelines may help determine whether the perceived Link_Formatting shortcomings should be addressed.	<ul style="list-style-type: none">• Links-Related Usability Guidelines at usability.gov• Navigation-Related Usability Guidelines at usability.gov

After you have looked at each of the evaluation results and the accompanying usability guidelines, **please complete the accompanying Web-based survey** which constitutes the second and final part of my survey study. The **goal** of this survey is to learn about whether you think the **kind of results** generated by an automated Web usability evaluation tool is helpful to you for designing Web sites.

- Please go to <http://www.web4success.org/survey/index.php?sid=2&token=6427763175> to start the survey.

If you have any questions, please contact Saun Shewanown at saun@uga.edu or 404-437-6356.

Figure E.9. Sample link results summary.

Usability Evaluation Results Summary

<http://www.sample.com>

Page

[How to Read the Evaluation Results](#)

Page_Title	Evaluation Results	Page-Related Usability Guidelines
Page Title refers to the use of "descriptive and different title on each page" (Koyani, Bailey, & Nall, 2004, p. 76) on a Web site.	Good The results suggest that Page_Title aspects of http://www.sample.com are good and comparable with other similar sites. You may consider consulting the Page related usability guidelines for a few Page_Title guidelines that may have been overlooked.	<ul style="list-style-type: none">• Page Title-Related Usability Guidelines at usability.gov• Page Layout-Related Usability Guidelines at usability.gov• Page Length-Related Usability Guidelines at usability.gov
Page_Formatting Page Formatting refers to the "color use, fonts, page size, used of interactive elements, page style control, and so on" (Ivory, 2003, p. 71) used on a Web site.	Excellent The results suggest that Page_Formatting aspects of http://www.sample.com are significantly better than other similar sites. Feel free to consult the Page related usability guidelines for helpful information about Page_Formatting usability.	

After you have looked at each of the evaluation results and the accompanying usability guidelines, **please complete the accompanying Web-based survey** which constitutes the second and final part of my survey study. The **goal** of this survey is to learn about whether you think the **kind of results** generated by an automated Web usability evaluation tool is helpful to you for designing Web sites.

- Please go to <http://www.web4success.org/survey/index.php?sid=2&token=6427763175> to start the survey.

If you have any questions, please contact Saun Shewanown at saun@uga.edu or 404-437-6356.

Figure E.10. Sample page results summary.

Reading the Evaluation Results for Each Category

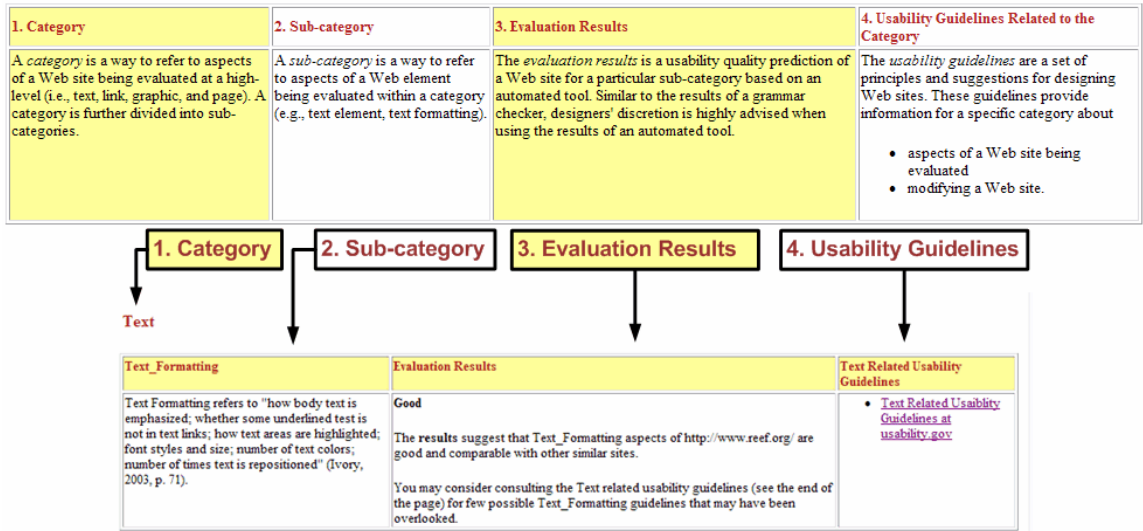


Figure E.11. A description of how to read the evaluation results summary.

Dear John,

Thank you for completing the first survey of my study. I have created a summary of the usability evaluation results for <http://www.sample.com>. These results were produced using an empirical-based automated usability evaluation tool. You are invited to participate in the second Web-based survey regarding your perceptions of how useful these results of an automated usability evaluation tool are in identifying Web site usability issues and improving Web site usability quality.

This second survey is also part of my dissertation study, entitled "Investigating Usability of Web-Based Learning Sites Using an Empirical Automated Usability Evaluation Approach and the Perceptions of Site Builders". My dissertation is under the supervision of Dr. Thomas C. Reeves (<http://it.coe.uga.edu/~treeves>).

This invitation is being sent to people who have completed the first Web-based survey of this study. Your identity will not be associated with the findings nor will it be released. This study has been approved by the Human Subjects Office of the University of Georgia (<http://www.ovpr.uga.edu/hso>).

Because your time is valuable, I limited this Web-based survey to only 4 items requesting your degree of agreement or disagreement. We would greatly appreciate it if you would complete the surveys by THURSDAY, NOVEMBER 10 at 3 P.M. This will allow us to meet the required semester deadlines. (Of course, we are still interested in your responses should you miss the deadline, so please complete the survey and forward any responses after the date above as well.)

I hope you will decide to contribute a small amount of your time to this study that ultimately could lead to the enhancement of Web-based learning. As an extra incentive, you will have a chance to win a \$100 gift certificate from www.amazon.com by completing in both surveys. University of Georgia employees will be excluded from this extra incentive.

Please go to <http://www.web4success.org/survey/results/results.php?selectedID=test01> to view usability evaluation results of <http://www.sample.com>.

After you have looked at each of the evaluation results, please complete the accompanying Web-based survey which constitutes the second and final part of my survey study. The link to the survey is located near the bottom of the evaluation results page.

I would be pleased to answer any additional questions you may have about this study. Thank you for your contribution to this study.

Sincerely,

Saun Shewanown
Doctoral Candidate
Instructional Technology, The University of Georgia
saun@uga.edu, 404-437-6356

Figure E.12. The invitation for the second survey instrument.

A Survey Study (Part Two): The Usefulness of the Results of an Automated Usability Evaluation Tool

Welcome,

Thank you very much for participating in my survey study to gain a better understanding of the awareness, perceptions and practice in evaluating usability of online learning resources such as the one with which you are associated.

You are now invited to participate in the second survey of a research study titled "Investigating Usability of Web-Based Learning Sites Using an Empirical Automated Usability Evaluation Approach and the Perceptions of Site Builders" conducted by Saun Shewanown, Department of Educational Psychology and Instructional Technology, University of Georgia, 404-437-6356 under the direction of Dr. Thomas C. Reeves, Department of Educational Psychology and Instructional Technology, University of Georgia, 603D Aderhold Hall, Athens, GA 30602-7144.

The purpose of this second survey is to understand, from your perspective, the usefulness of the results of an automated usability evaluation tool in identifying Web site usability issues and improving your Web-based learning site's usability quality. In the following screens, you will be asked to respond to a brief (4 question) survey.

Completion of the survey is expected to take a maximum of 5 minutes. Please note that Internet communications are insecure and there is a limit to the confidentiality that can be guaranteed due to the technology itself. However, once I receive the completed surveys, I will store the data in a locked cabinet in my office and will destroy them and any names and contact information that I have by June, 2006. Any information that is obtained in connection with this study and that can be identified with you will remain confidential except as required by law. If you are not comfortable with the level of confidentiality provided by the Internet, please feel free to print out a copy of the survey, fill it out by hand, and mail it to me at the address given below, with no return address on the envelope.

Your participation in this study is completely voluntary. You may withdraw at any time without penalty, or skip any questions you feel uncomfortable answering. Closing the survey window will erase your answers without submitting them. Additionally, you will be given a choice of submitting or discarding your responses at the end of the survey. If you have any questions do not hesitate to ask now or at a later date. You may contact Saun Shewanown at 404-437-6356 or saun@uga.edu.

Thank you for the invaluable help that you are providing by participating in this research study.

Sincerely,

Saun Shewanown
EPIT Department, University of Georgia
604 Aderhold Hall, Athens, Georgia 30602
saun@uga.edu, 404-437-6356

By completing the survey you are agreeing to participate in the research.
Please begin the survey **below**.
Click on the **Submit** button at the bottom of this page when you are done.

Figure E.13. The informed consent form for the second survey instrument.

I am interested in your perceptions of the summary of the usability evaluation results of your Web site that you have received from me recently. Please check your agreement with each of the following statements using the scale below.
 1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

	1	2	3	4	5	N/A
#1. The summary of the usability results regarding my Web site was easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#2. The results accurately reflect the usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#3. The results accurately reflect usability issues of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#4. The results would be helpful for improving usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Figure E.14. The second survey instrument addressing research question #3.

Appendix F: The Pilot Study Results Histograms

The First Survey Instrument: Addressing Research Questions #2.1 and #2.2

This section provides the figures that illustrate the *central tendency*, *variability*, *size*, and *distribution* of questions #1 through #6 of the first survey instrument in the pilot study. This section of the first survey (see Figure F.1) addresses research question #2.1 and #2.2.

Please check your **level of expertise** with the following [usability examination methods](#) using a scale where 1 = "Not familiar with it" to 5 = "expert in its application."

#1. Usability Testing: User(s) perform predetermined tasks while being observed and recorded. (

#2. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria. (

#3. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality. (

Please check your **use** of the following [usability examination methods](#) using a scale where 1 = "Never" to 5 = "always."

#4. Usability Testing: User(s) perform predetermined tasks while being observed and recorded. (

#5. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria. (

#6. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality. (

Figure F.1. Questions #1 through #6 of the first survey instrument which addressed research questions #2.1 and #2.2.

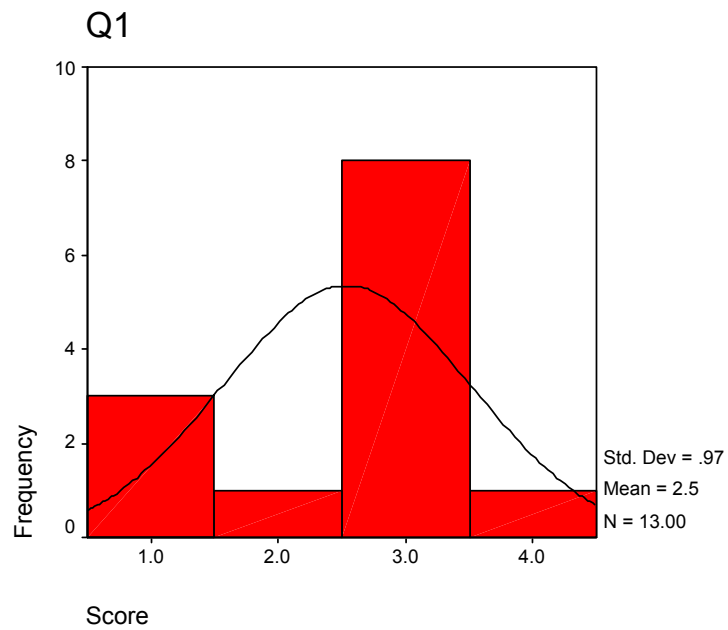


Figure F.2. The central tendency, variability, size, and distribution of question #1 of the first survey instrument.

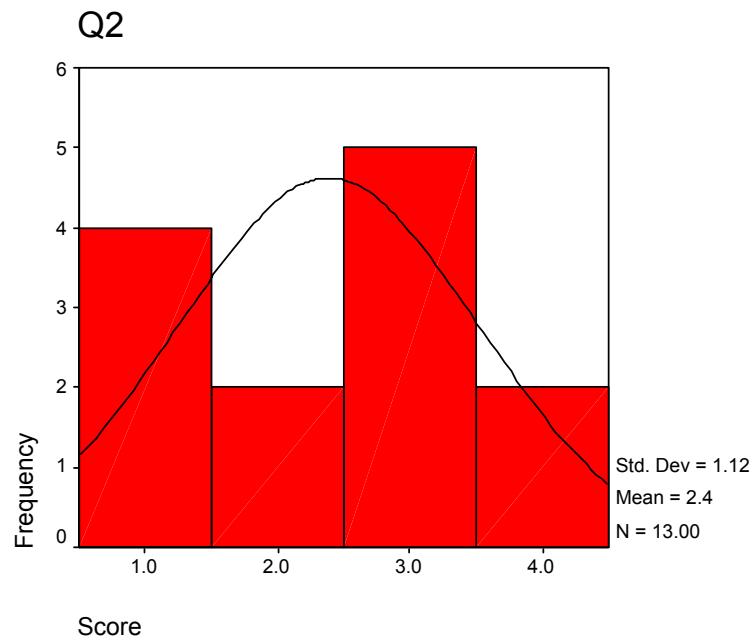


Figure F.3. The central tendency, variability, size, and distribution of question #2 of the first survey instrument.

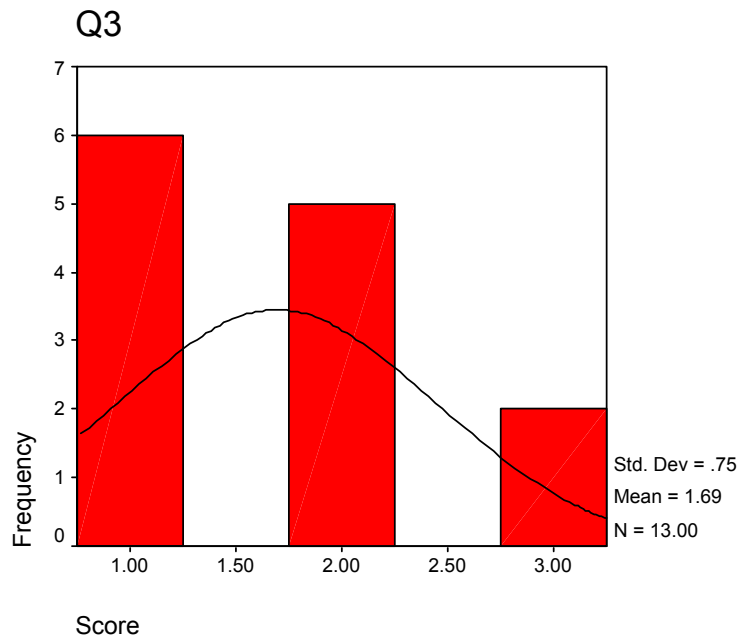


Figure F.4. The central tendency, variability, size, and distribution of question #3 of the first survey instrument.

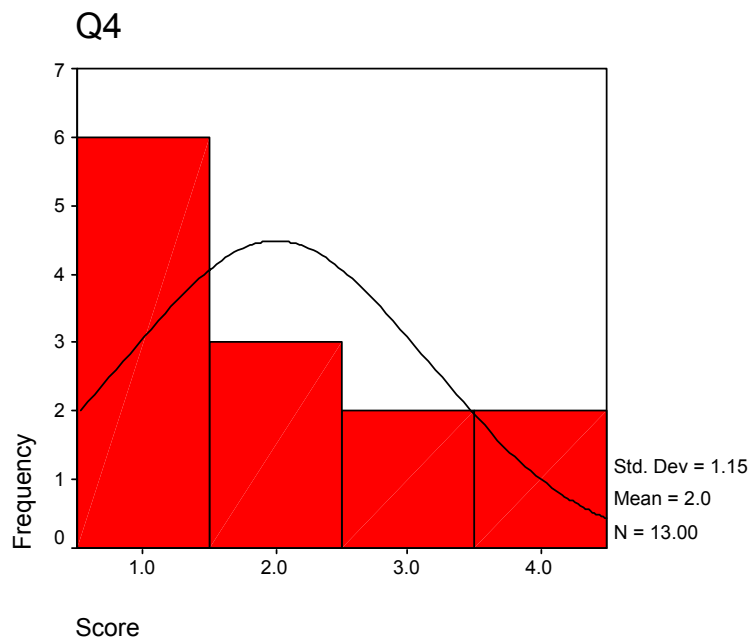


Figure F.5. The central tendency, variability, size, and distribution of question #4 of the first survey instrument.

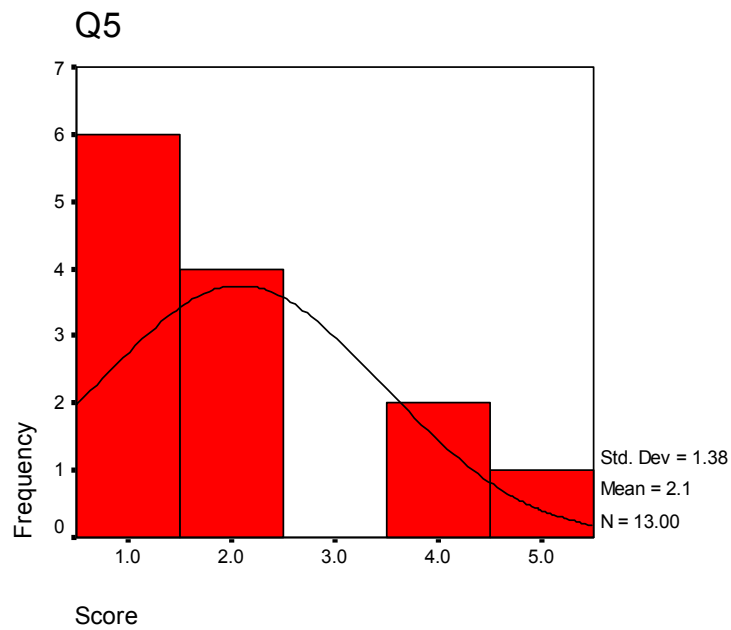


Figure F.6. The central tendency, variability, size, and distribution of question #5 of the first survey instrument.

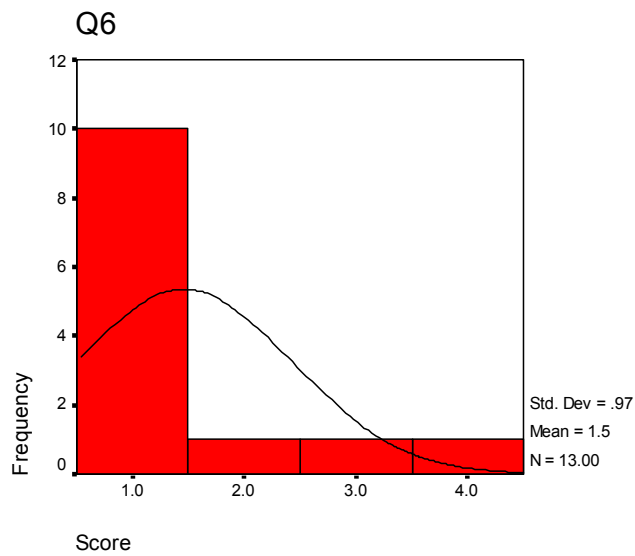


Figure F.7. The central tendency, variability, size, and distribution of question #6 of the first survey instrument.

The First Survey Instrument: Addressing Research Questions #2.3

This section provides the figures that illustrate the *central tendency*, *variability*, *size*, and *distribution* of questions #7 through #12 of the first survey instrument in the pilot study. Questions #7 through #12 of the first survey instrument (see Figure F.8) addressed research question #2.3.

Please check your agreement with each of the following statements using the scale below.
1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

#7. I am interested in further improvement of the usability of my Web site.

#8. Usability is an essential feature of online learning resources.

#9. Usability is usually overlooked in the design of online learning resources.

#10. Personally, I would like to know more about usability in the context of online learning resources.

#11. It would be very helpful to automate the process of finding usability problems in online learning resources.

#12. Personally, I am skeptical about the value of automating the process of finding usability problems in online learning resources.

Figure F.8. Questions #7 through #12 of the first survey instrument which addressed research question #2.3.

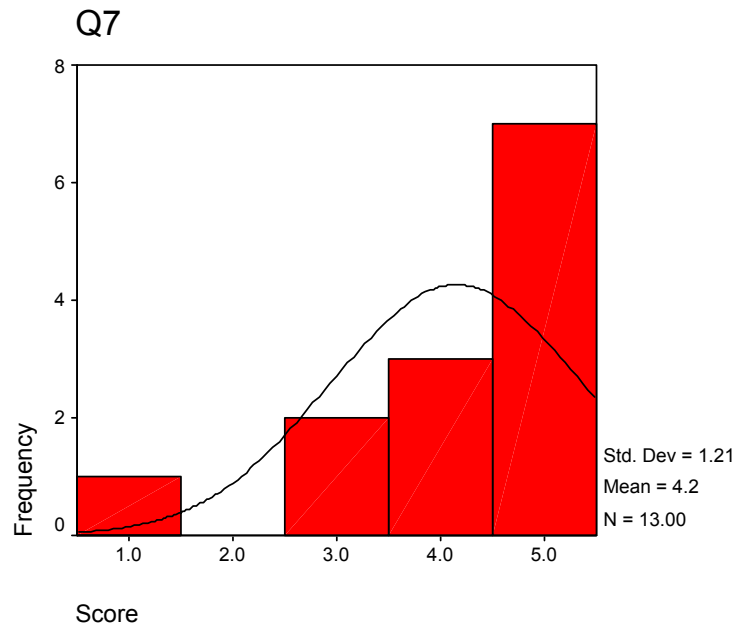


Figure F.9. The central tendency, variability, size, and distribution of question #7 of the first survey instrument.

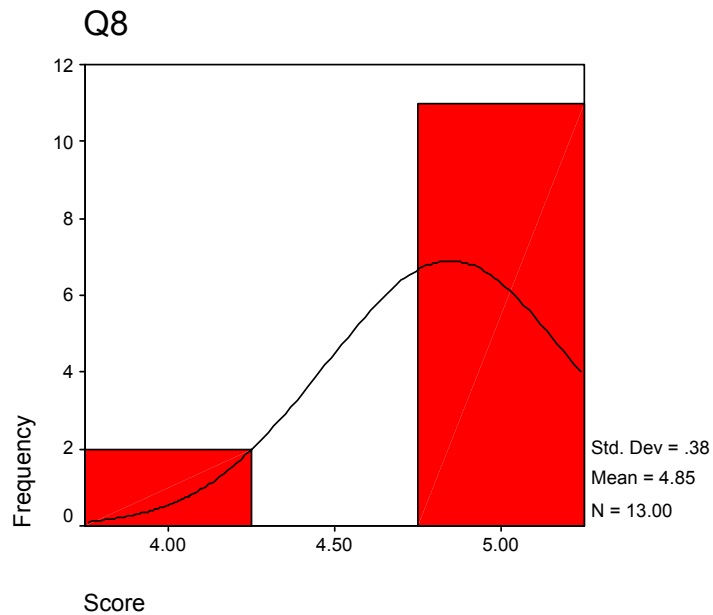


Figure F.10. The central tendency, variability, size, and distribution of question #8 of the first survey instrument.

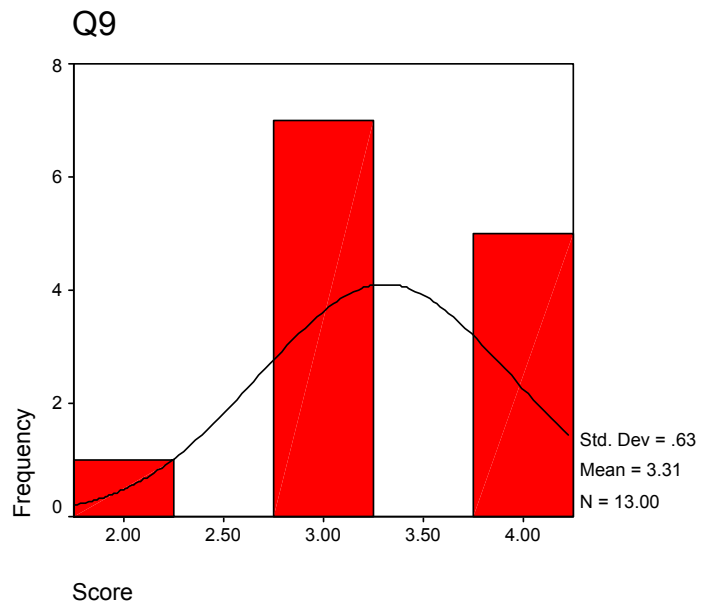


Figure F.11. The central tendency, variability, size, and distribution of question #9 of the first survey instrument.

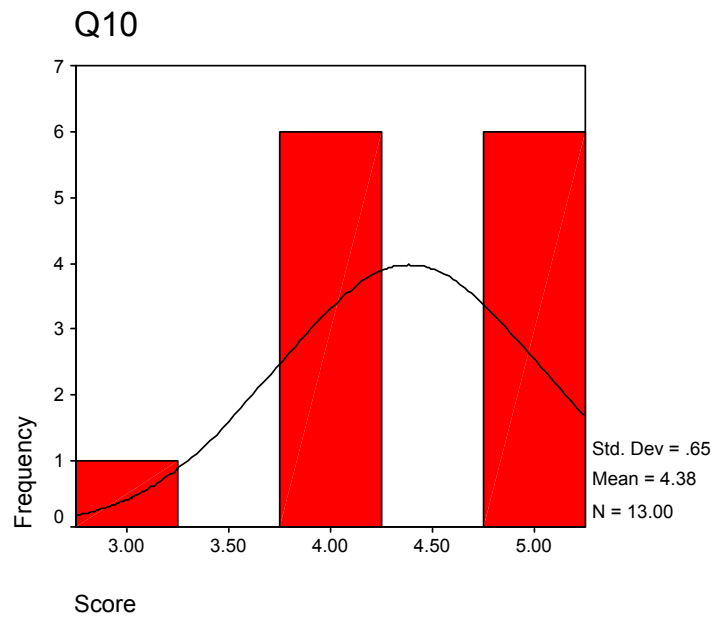


Figure F.12. The central tendency, variability, size, and distribution of question #10 of the first survey instrument.

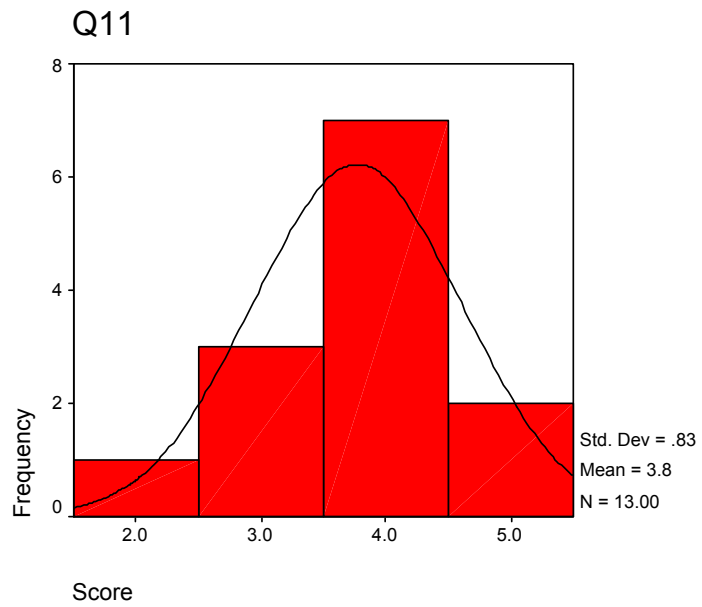


Figure F.13. The central tendency, variability, size, and distribution of question #11 of the first survey instrument.

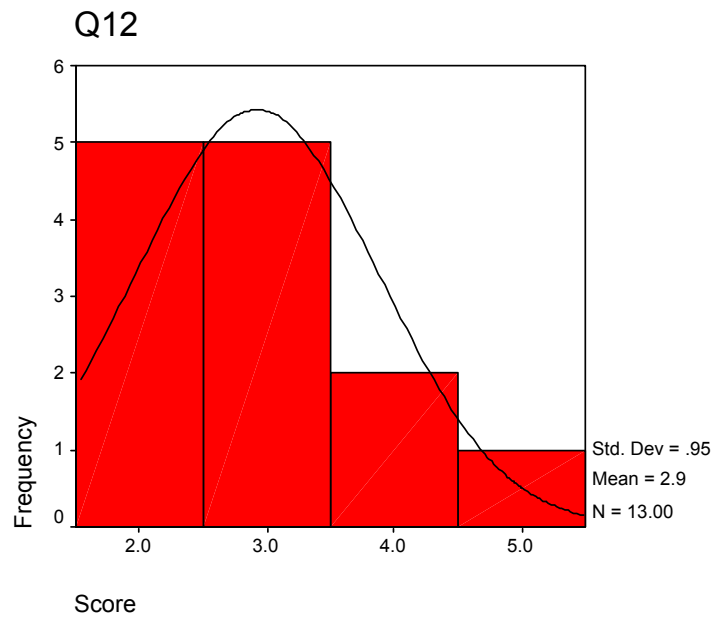


Figure F.14. The central tendency, variability, size, and distribution of question #12 of the first survey instrument.

The Second Survey Instrument: Addressing Research Question #3

This section provides the figures that illustrate the *central tendency*, *variability*, *size*, and *distribution* of questions #1 through #4 of the second survey instrument in the pilot study. The second survey (see Figure F.15) addresses research question #3.

I am interested in your perceptions of the summary of the usability evaluation results of your Web site that you have received from me recently. Please check your agreement with each of the following statements using the scale below.
1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

	1	2	3	4	5	N/A
#1. The summary of the usability results regarding my Web site was easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#2. The results accurately reflect the usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#3. The results accurately reflect usability issues of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#4. The results would be helpful for improving usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Figure F.15. Questions #1 through #4 of the second survey instrument.

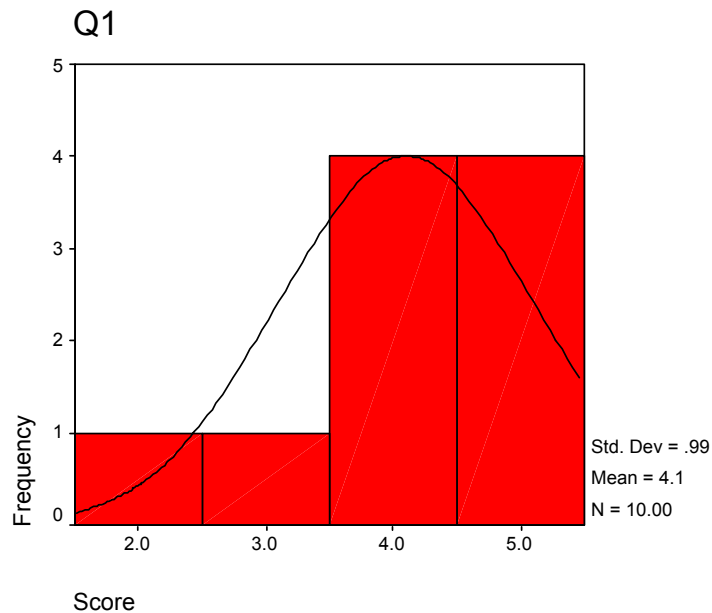


Figure F.16. The central tendency, variability, size, and distribution of question #1 of the second survey instrument.

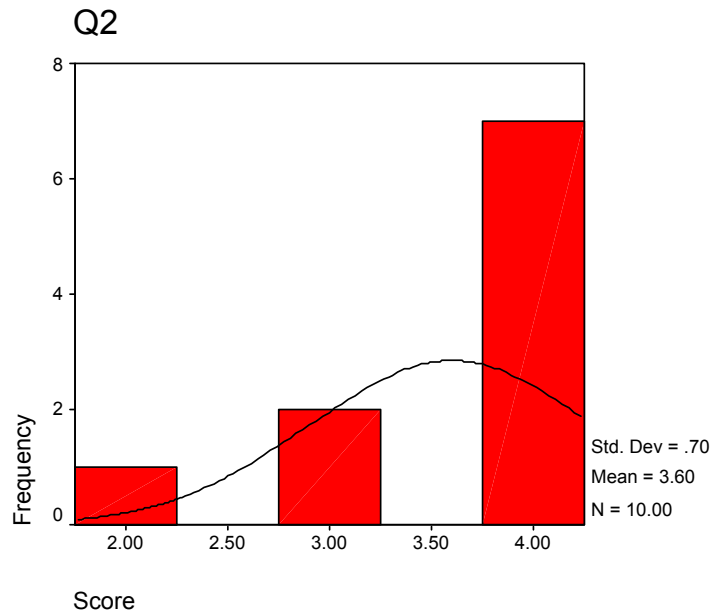


Figure F.17. The central tendency, variability, size, and distribution of question #2 of the second survey instrument.

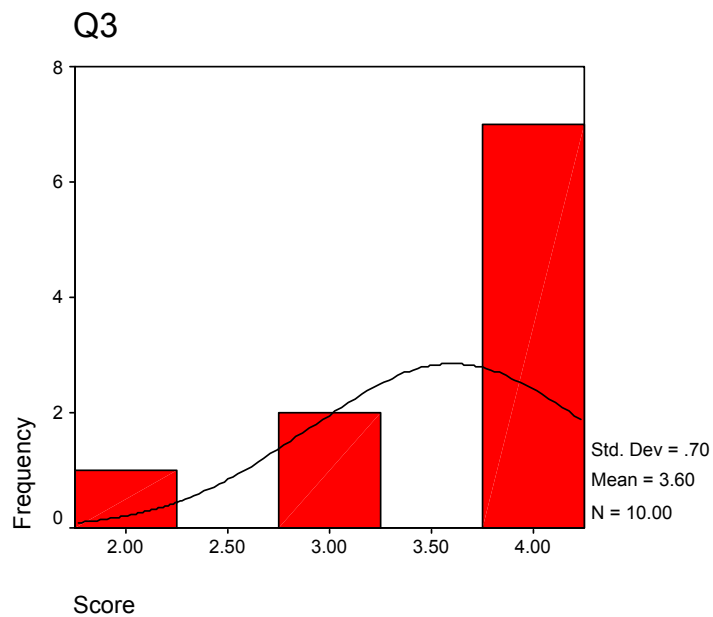


Figure F.18. The central tendency, variability, size, and distribution of question #3 of the second survey instrument.

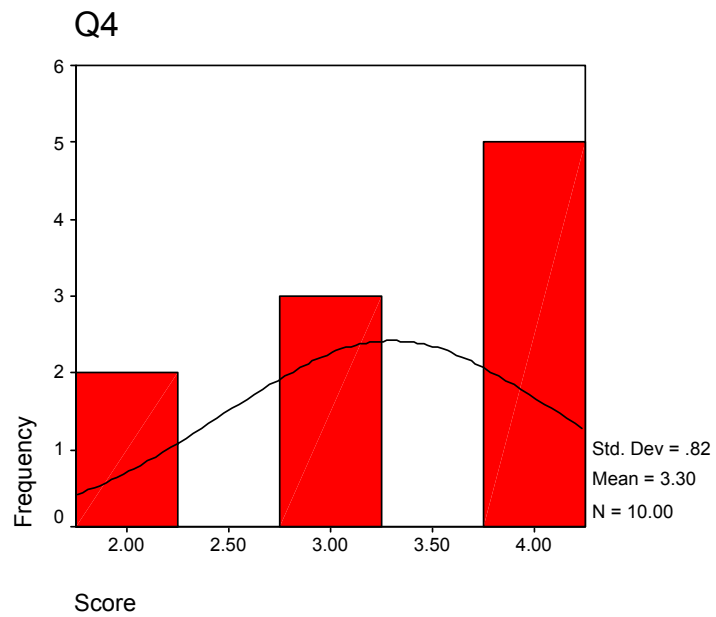


Figure F.19. The central tendency, variability, size, and distribution of question #4 of the second survey instrument.

Appendix G: The Results Chapter Histograms

The First Survey Instrument: Addressing Research Questions #2.1 and #2.2

This section provides the figures that illustrate the *central tendency*, *variability*, *size*, and *distribution* of questions #1 through #6 of the first survey instrument of the survey study. These section of the first survey (see Figure G.1) addresses research questions #2.1 and #2.2.

Please check your **level of expertise** with the following [usability examination methods](#) using a scale where 1 = "Not familiar with it" to 5 = "expert in its application."

#1. Usability Testing: User(s) perform predetermined tasks while being observed and recorded.

#2. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.

#3. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.

Please check your **use** of the following [usability examination methods](#) using a scale where 1 = "Never" to 5 = "always."

#4. Usability Testing: User(s) perform predetermined tasks while being observed and recorded.

#5. Usability Inspection or Heuristic Evaluation: A process whereby an evaluator examines the usability aspects of a Web site with respect to the site's conformance to a set of guidelines or criteria.

#6. Automated Usability Evaluation: Fully or partly automated software tools or methods for analyzing Web site user interfaces to determine the usability quality.

Figure G.1. Questions #1 through #6 of the first survey instrument which addressed research questions #2.1 and #2.2.

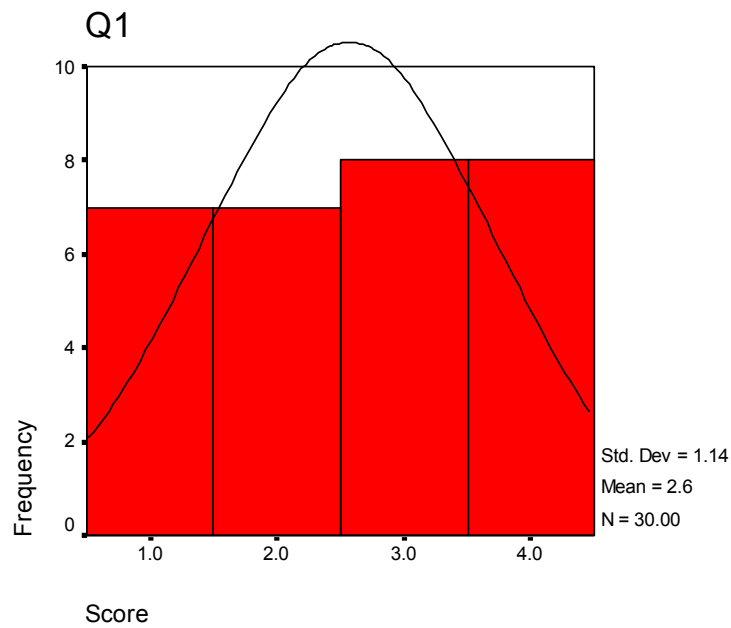


Figure G.2. The central tendency, variability, size, and distribution of question #1 of the first survey instrument.

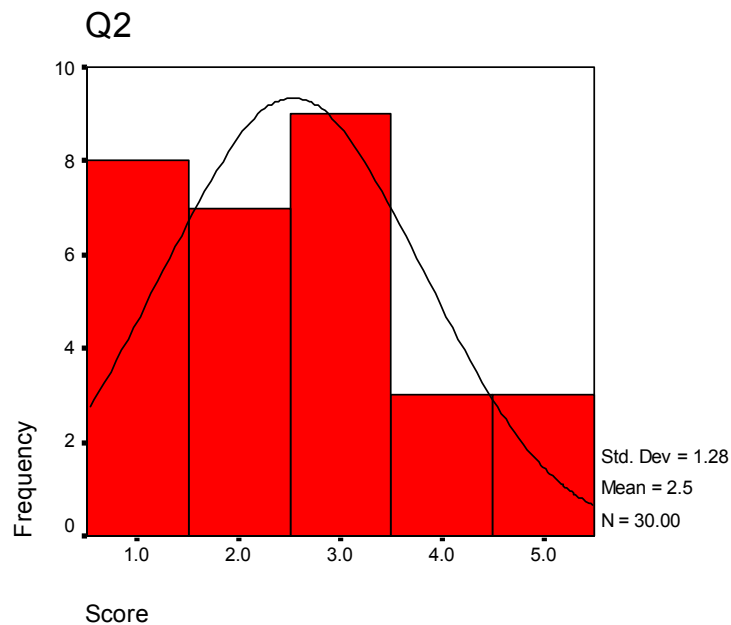


Figure G.3. The central tendency, variability, size, and distribution of question #2 of the first survey instrument.

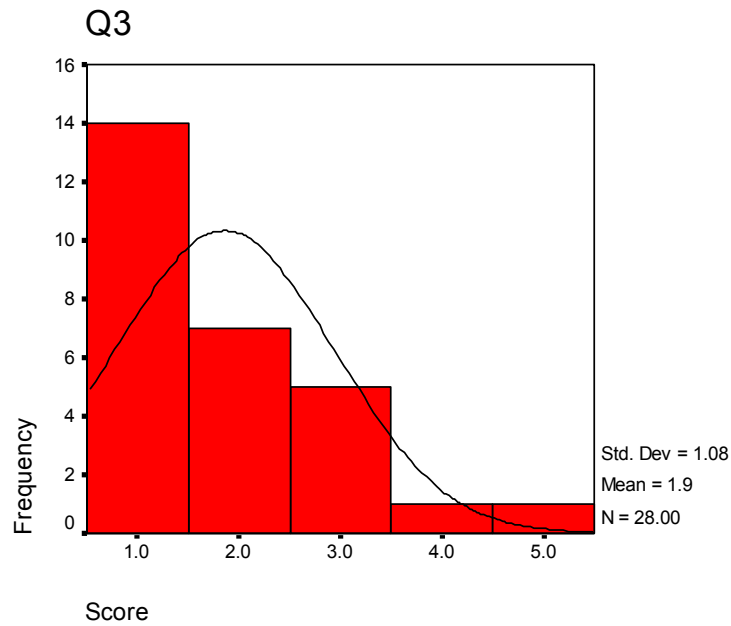


Figure G.4. The central tendency, variability, size, and distribution of question #3 of the first survey instrument.

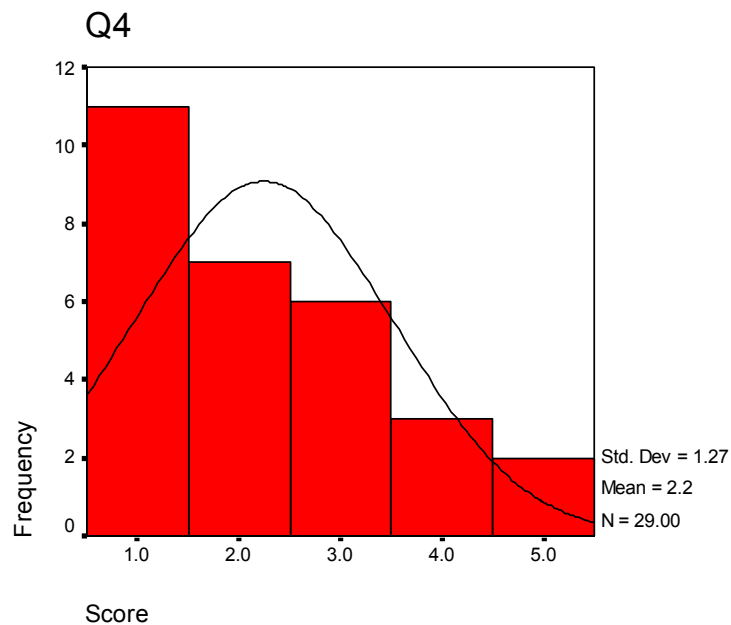


Figure G.5. The central tendency, variability, size, and distribution of question #4 of the first survey instrument.

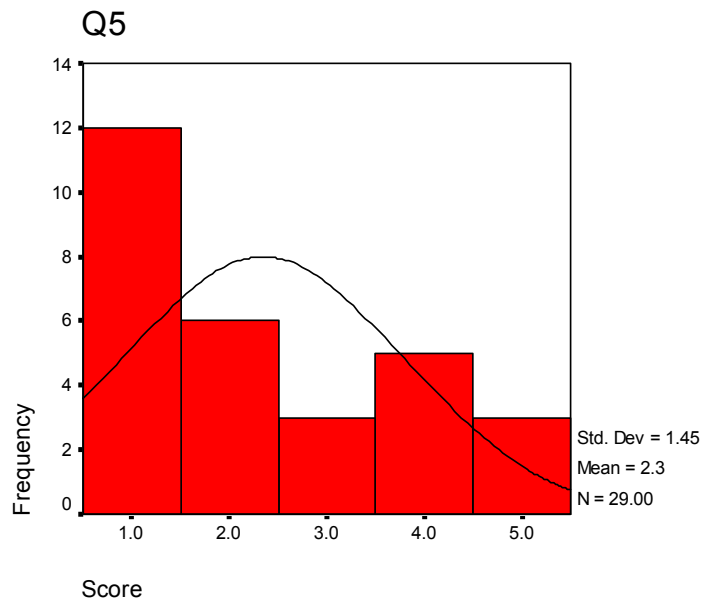


Figure G.6. The central tendency, variability, size, and distribution of question #5 of the first survey instrument.

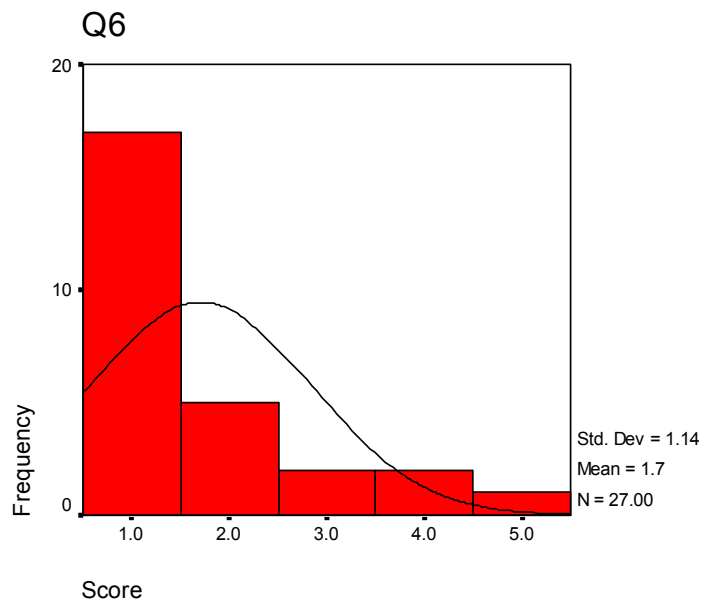


Figure G.7. The central tendency, variability, size, and distribution of question #6 of the first survey instrument.

The First Survey Instrument: Addressing Research Question #2.3

This section provides the figures that illustrate the *central tendency*, *variability*, *size*, and *distribution* of questions #7 through #12 of the first survey instrument. This section of the first survey (see Figure G.8) addresses research question #2.3.

Please check your agreement with each of the following statements using the scale below.
1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

#7. I am interested in further improvement of the usability of my Web site.

#8. Usability is an essential feature of online learning resources.

#9. Usability is usually overlooked in the design of online learning resources.

#10. Personally, I would like to know more about usability in the context of online learning resources.

#11. It would be very helpful to automate the process of finding usability problems in online learning resources.

#12. Personally, I am skeptical about the value of automating the process of finding usability problems in online learning resources.

Figure G.8. Questions #7 through #12 of the first survey instrument which addressed research questions #2.3.

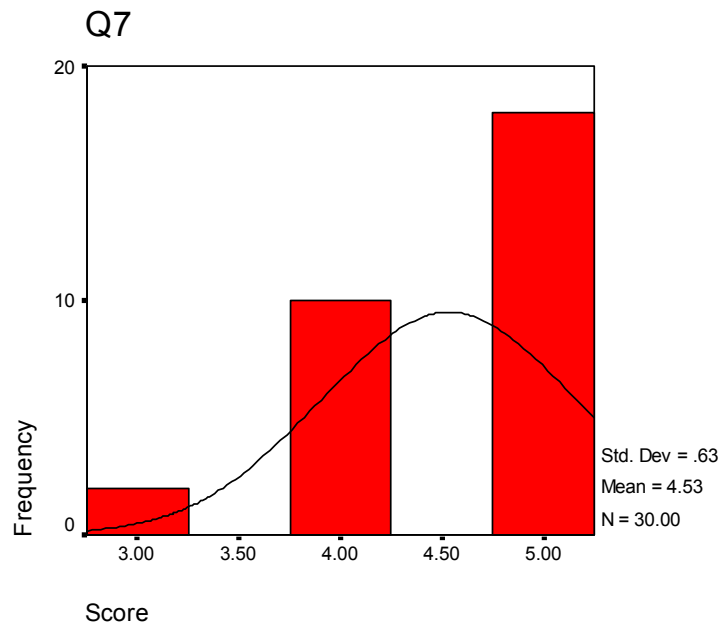


Figure G.9. The central tendency, variability, size, and distribution of question #7 of the first survey instrument.

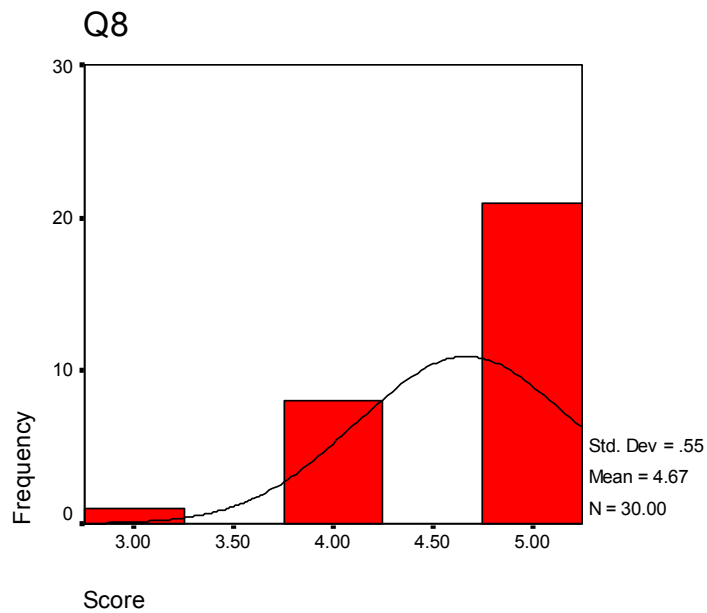


Figure G.10. The central tendency, variability, size, and distribution of question #8 of the first survey instrument.

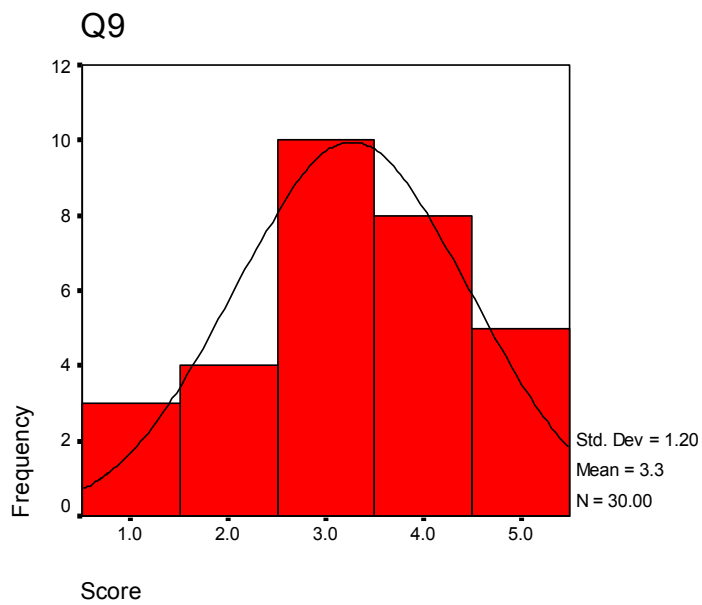


Figure G.11. The central tendency, variability, size, and distribution of question #9 of the first survey instrument.

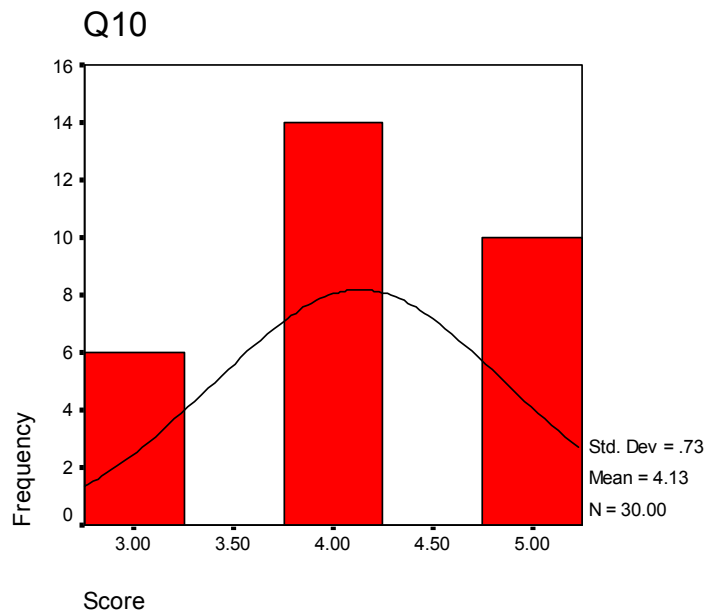


Figure G.12. The central tendency, variability, size, and distribution of question #10 of the first survey instrument.

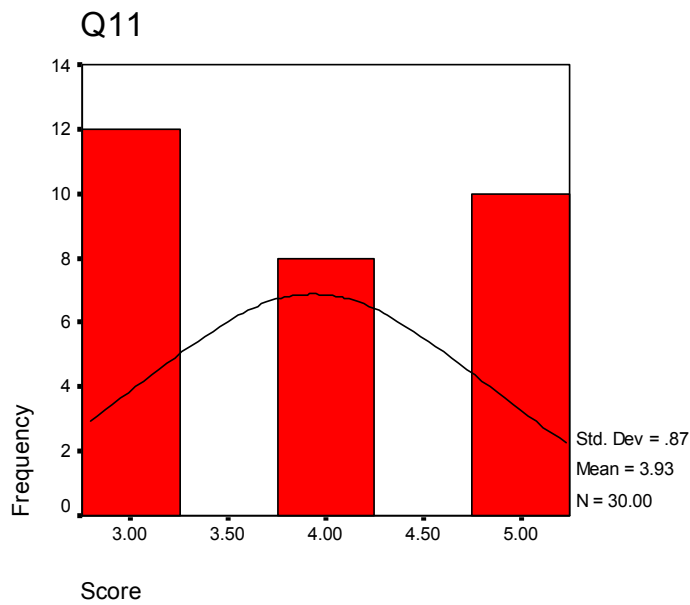


Figure G.13. The central tendency, variability, size, and distribution of question #11 of the first survey instrument.

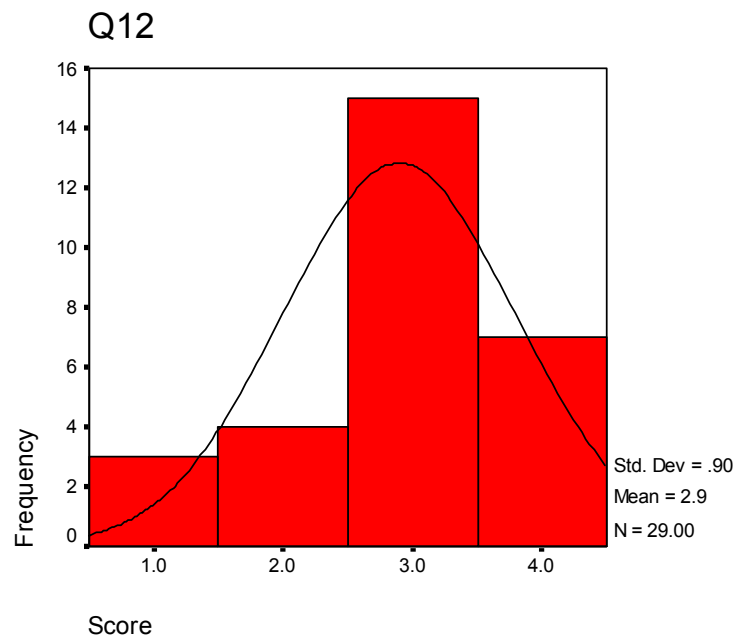


Figure G.14. The central tendency, variability, size, and distribution of question #12 of the first survey instrument.

The Second Survey Instrument: Addressing Research Question #3

This section provides the figures that illustrate the *central tendency*, *variability*, *size*, and *distribution* of questions #1 through #4 of the second survey instrument. The second survey (see Figure G.15) addresses research question #3.

I am interested in your perceptions of the summary of the usability evaluation results of your Web site that you have received from me recently. Please check your agreement with each of the following statements using the scale below.
1 = Strongly Disagree 2 = Disagree 3 = Neither Agree nor Disagree 4 = Agree 5 = Strongly Agree N/A = Not Applicable

	1	2	3	4	5	N/A
#1. The summary of the usability results regarding my Web site was easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#2. The results accurately reflect the usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#3. The results accurately reflect usability issues of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
#4. The results would be helpful for improving usability quality of my Web site.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Figure G.15. Questions #1 through #4 of the second survey instrument.

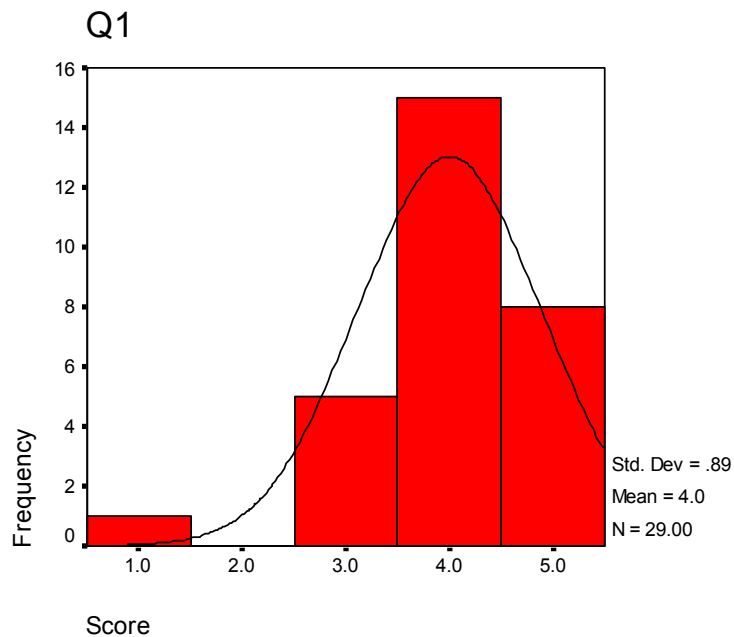


Figure G.16. The central tendency, variability, size, and distribution of question #1 of the second survey instrument.

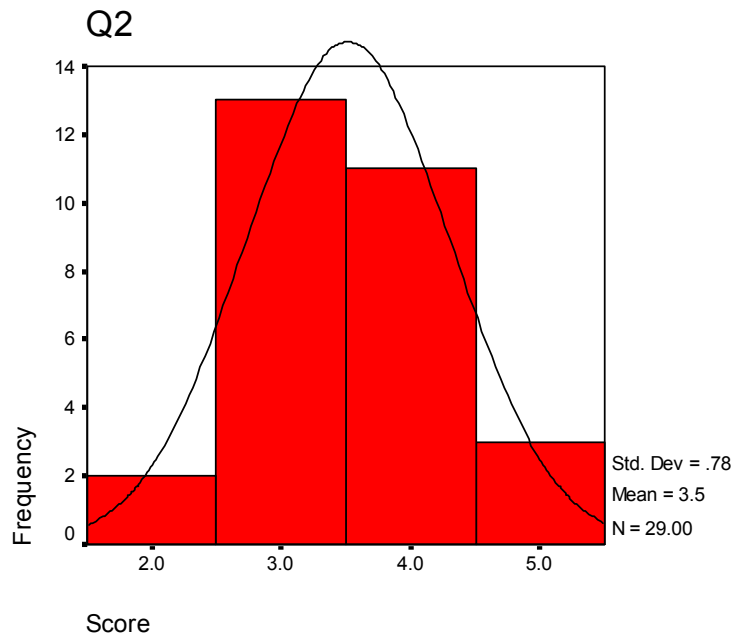


Figure G.17. The central tendency, variability, size, and distribution of question #2 of the second survey instrument.

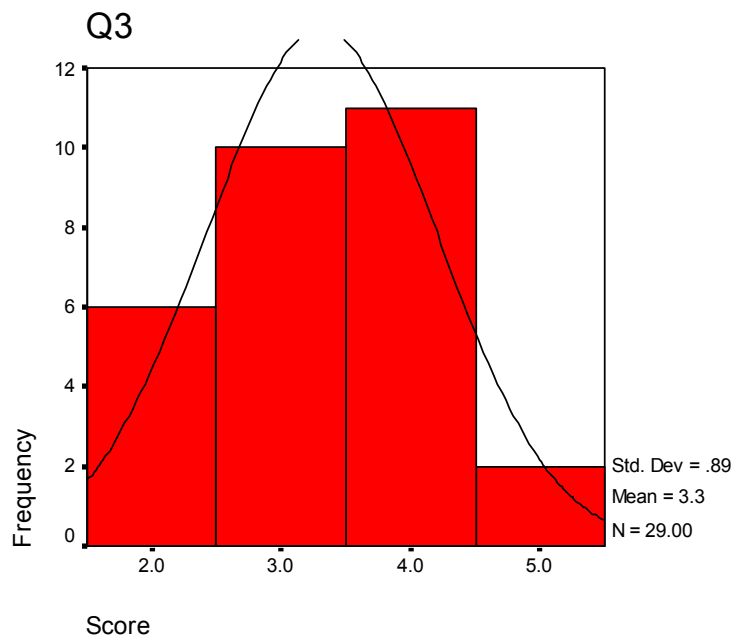


Figure G.18. The central tendency, variability, size, and distribution of question #3 of the second survey instrument.

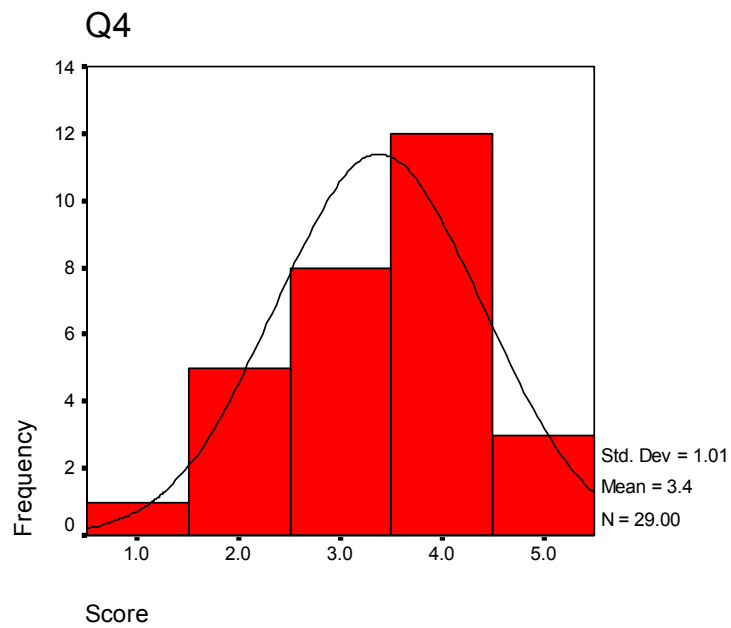


Figure G.19. The central tendency, variability, size, and distribution of question #4 of the second survey instrument.

Appendix H: Diagram Symbols

