USABILITY OF THE DIGITAL LIBRARY:
AN EVALUATION MODEL

by

JUDY H. JENG

A Dissertation submitted to the
Graduate School – New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Communication, Information, and Library Studies
written under the direction of
Professor Tefko Saracevic
and approved by

New Brunswick, New Jersey
January, 2006
ABSTRACT OF THE DISSERTATION

Usability of the Digital Library:
An Evaluation Model
By JUDY H. JENG

Dissertation Director: Professor Tefko Saracevic

The main research goal of this dissertation is to develop a model and a suite of instruments to evaluate the usability of academic digital libraries. Usability in this study was examined from the perspectives of effectiveness, efficiency, satisfaction, and learnability. The model that was developed considers both quantifying elements such as accuracy rate, time, and number of steps required to complete a task, and the subjective criterion, satisfaction, which is further broken down into the areas of ease of use, organization of information, terminology and labeling, visual appearance, content, and mistake recovery. The model was tested at two academic digital libraries. It is a cross-institutional usability study. The study recruited a total of forty-one subjects and was divided into two stages to confirm findings.

The study examined the interlocking relationships among effectiveness, efficiency, and satisfaction. It found that the relationships are statistically significant. The longer the time it took for a subject to perform a task, the more steps involved. The greater the number of steps involved in completing a task, the lower the satisfaction expressed by the subject. The longer the time spent on completing a task, the lower the satisfaction expressed by the subject. However, it should be noted that each criterion has its own emphasis and should be measured separately.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The research uncovered the user's criteria for *ease of use, organization of information, terminology, visual attractiveness,* and *mistake recovery.* The findings are helpful in enhancing the usability of the digital library.

The issues of “user lostness” and navigation disorientation were examined, and the causes were identified.

The study also examined the *click cost* issue. It found that 73% of the participants expected the click(s) to lead them eventually to the correct answer. Each click in the digital library should bring users closer to the answer.

The study established that the demographic factors (gender, age, status, academic major, ethnic background, years at the institution, and frequency of using library Web sites) do not have a statistically significant impact on performance (i.e., effectiveness and efficiency). However, different attitudes among ethnic groups were found in rating satisfaction. Future research should recruit more subjects in each ethnic group to examine this cross-culture usability issue further.
ACKNOWLEDGMENTS

I sincerely want to thank the following people for their assistance to me throughout this endeavor: Tefko Saracevic, Nick Belkin, Colleen Cool, Dan O’Connor, and Marija Dalbello.
DEDICATION

This dissertation is dedicated with love to my parents, my husband, and my daughters. Without my family's love, understanding, encouragement, and support, I could not have accomplished this.
# TABLE OF CONTENTS

Abstract ii  
Acknowledgments iv  
Dedication v  
Table of Contents vi  
List of Tables x  
List of Figures xiii

CHAPTER 1  INTRODUCTION 1  
1.1 Background 1  
1.2 Problem Statement 2  
1.3 Research Goals 3  
1.4 Organization of This Dissertation 3

CHAPTER 2  LITERATURE REVIEW 5  
2.1 Definitions of Digital Library 5  
2.2 Dimensions of Usability 10  
2.3 Techniques of Usability Evaluation 20  
2.4 Selected Usability Tests of Digital Libraries 24  
2.5 Synthesis of Findings from Digital Library Usability Tests 38

CHAPTER 3  USABILITY EVALUATION MODEL 45

CHAPTER 4  METHODS OF DATA COLLECTION 51  
4.1 Sites 51  
4.2 Stages 51  
4.3 Subjects 52  
4.4 Tasks 52  
4.5 Instruments 53  
4.6 Setting 54  
4.7 Procedure 55  
4.8 Pilot Study 55

CHAPTER 5  DATA ANALYSIS 56  
5.1 Demographic Distribution 56  
5.2 Effectiveness 59  
5.3 Efficiency 64  
5.4 Learnability 69  
5.5 Satisfaction 73  
5.5.1 Overall Satisfaction 73  
5.5.2 Ease of Use 77  
5.5.3 Organization of Information 81  
5.5.4 Terminology 86  
5.5.5 Attractiveness 88  
5.5.6 Mistake Recovery 92
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.</td>
<td>Attributes of Usability</td>
<td>17</td>
</tr>
<tr>
<td>Table 2.</td>
<td>Usability Assessment at Academic Sites</td>
<td>34</td>
</tr>
<tr>
<td>Table 3.</td>
<td>Factors of the Evaluation Model</td>
<td>49</td>
</tr>
<tr>
<td>Table 4.</td>
<td>Participants Demographic Data</td>
<td>58</td>
</tr>
<tr>
<td>Table 5.</td>
<td>Effectiveness of Rutgers Subjects Searching Rutgers Site and Queens Subjects Searching Queens Site</td>
<td>59</td>
</tr>
<tr>
<td>Table 6.</td>
<td>Effectiveness of All Subjects Searching Rutgers and Queens Sites</td>
<td>60</td>
</tr>
<tr>
<td>Table 7.</td>
<td>Time Used by Subjects to Search Their Own Institutions’ Sites</td>
<td>65</td>
</tr>
<tr>
<td>Table 8.</td>
<td>Time Used by All Subjects in Stages One and Two on Both Sites</td>
<td>65</td>
</tr>
<tr>
<td>Table 9.</td>
<td>Number of Steps Taken by Rutgers Subjects to Search Rutgers Site and Queens Subjects to Search Queens Site</td>
<td>66</td>
</tr>
<tr>
<td>Table 10.</td>
<td>Number of Steps by All Subjects in Stages One and Two on Both Sites</td>
<td>67</td>
</tr>
<tr>
<td>Table 11.</td>
<td>Learnability Comparison by Correctness of Answers in Stage One</td>
<td>70</td>
</tr>
<tr>
<td>Table 12.</td>
<td>Learnability Comparison by Correctness of Answers in Stage Two</td>
<td>70</td>
</tr>
<tr>
<td>Table 13.</td>
<td>Learnability by Time in Stage One</td>
<td>71</td>
</tr>
<tr>
<td>Table 14.</td>
<td>Learnability by Time in Stage Two</td>
<td>72</td>
</tr>
<tr>
<td>Table 15.</td>
<td>Learnability of Both Sites in Stage One</td>
<td>72</td>
</tr>
<tr>
<td>Table 16.</td>
<td>Learnability of Both Sites in Stage Two</td>
<td>73</td>
</tr>
<tr>
<td>Table 17.</td>
<td>Satisfaction Rating by Question</td>
<td>74</td>
</tr>
<tr>
<td>Table 18.</td>
<td>Satisfaction by Site</td>
<td>74</td>
</tr>
<tr>
<td>Table 19.</td>
<td>Subjects’ Comments on Overall Satisfaction</td>
<td>77</td>
</tr>
<tr>
<td>Table 20.</td>
<td>Ratings of Ease of Use</td>
<td>77</td>
</tr>
<tr>
<td>Table 21.</td>
<td>Subjects’ Comments on Ease of Use</td>
<td>81</td>
</tr>
<tr>
<td>Table 22.</td>
<td>Ratings of Organization of Information</td>
<td>82</td>
</tr>
<tr>
<td>Table 23.</td>
<td>Subjects' Comments on Organization of Information</td>
<td>85</td>
</tr>
<tr>
<td>Table 24.</td>
<td>Ratings of Terminology</td>
<td>86</td>
</tr>
<tr>
<td>Table 25.</td>
<td>Subjects' Comments on Terminology</td>
<td>88</td>
</tr>
<tr>
<td>Table 26.</td>
<td>Ratings of Attractiveness</td>
<td>89</td>
</tr>
<tr>
<td>Table 27.</td>
<td>Subjects' Comments on Visual Attractiveness</td>
<td>91</td>
</tr>
<tr>
<td>Table 28.</td>
<td>Ratings of Mistake Recovery</td>
<td>92</td>
</tr>
<tr>
<td>Table 29.</td>
<td>Subjects' Comments on Mistake Recovery</td>
<td>93</td>
</tr>
<tr>
<td>Table 30.</td>
<td>Ratings of Overall Reaction</td>
<td>94</td>
</tr>
<tr>
<td>Table 31.</td>
<td>Subjects' Comments on Overall Reaction</td>
<td>95</td>
</tr>
<tr>
<td>Table 32.</td>
<td>Subjects' Comments on Best Features</td>
<td>99</td>
</tr>
<tr>
<td>Table 33.</td>
<td>Subjects' Comments on Worst Features</td>
<td>103</td>
</tr>
<tr>
<td>Table 34.</td>
<td>Subjects' Comments on Desired Features</td>
<td>107</td>
</tr>
<tr>
<td>Table 35.</td>
<td>User Lostness</td>
<td>108</td>
</tr>
<tr>
<td>Table 36.</td>
<td>Ease of Navigation</td>
<td>109</td>
</tr>
<tr>
<td>Table 37.</td>
<td>Click Cost</td>
<td>111</td>
</tr>
<tr>
<td>Table 38.</td>
<td>Differences in Satisfaction Ranking between Correct or Incorrect Answers</td>
<td>113</td>
</tr>
<tr>
<td>Table 39.</td>
<td>Effectiveness and Steps</td>
<td>115</td>
</tr>
<tr>
<td>Table 40.</td>
<td>Effectiveness and Time</td>
<td>117</td>
</tr>
<tr>
<td>Table 41.</td>
<td>Differences among Ethnic Groups on Ease of Use on Queens Site</td>
<td>132</td>
</tr>
<tr>
<td>Table 42.</td>
<td>Differences among Ethnic Groups on Rating of Organization on Queens Site</td>
<td>134</td>
</tr>
<tr>
<td>Table 43.</td>
<td>Differences among Ethnic Groups on Rating of Organization on Both Sites Combined</td>
<td>135</td>
</tr>
</tbody>
</table>
Table 44. Differences among Ethnic Groups on Rating of Attractiveness on Queens Site 137
Table 45. Differences among Ethnic Groups on Rating of Attractiveness on Both Sites Combined 138
Table 46. Statistical Significance between Ethnic Background and Satisfaction 139
Table 47. Results of Effectiveness Test 157
Table 48. Results of Efficiency Test in Terms of Time 157
Table 49. Results of Efficiency Test in Terms of Steps 158
LIST OF FIGURES

Figure 1. The Four Principle Components in a Human-Machine System 15
Figure 2. A Model of Usability Evaluation for the Digital Library 48
Figure 3. Queens College Library Web Page in June 2004 61
Figure 4. Queens College Library Web Site in June 2004, Library Catalogs 62
Figure 5. Queens College Library Web Site in June 2004, E-Books 63
Figure 6. NetLibrary Logon from Queens College Library Site 64
Figure 7. Rutgers Online Catalog 76
Figure 8. Another Interface of Rutgers Online Catalog 79
Figure 9. Rutgers University Libraries Web Site in September 2004 97
Figure 10. Significant Differences in Steps Related to Correctness of Answers 116
Figure 11. Significant Differences in Time Related to Correctness of Answers 117
Figure 12. Interlocking Relationships among Effectiveness, Efficiency, and Satisfaction 150
Chapter 1

Introduction

This chapter provides the background of this dissertation research, its problem statement and research goals, and a description of the organization of the remaining chapters.

1.1 Background

Digital library development, since its inception in the 1990s, has made significant progress. Although there is still a long way to go before reaching its full potential, the digital library is “maturing” (Fox, 2002; Marcum, 2002). However, the evaluation of digital libraries has not kept pace. Saracevic (2005) pointed out that “Everything about digital libraries is explosive, except one thing: evaluation” (p. 1). He also found that “up to now there are more works that discuss evaluation than those that report evaluation” (p. 2). Saracevic (2000) also pointed out that several fundamental concepts remain to be clarified, such as: What is a digital library? What is there to evaluate? What are the criteria? How are they applied in evaluation? Why evaluate digital libraries in the first place? Borgman, Sølvberg, and Kovács (2002) have stated that the digital libraries research community needs large test beds, including collections and testing mechanisms, as a means to evaluate new concepts. There is also a need for benchmarks to provide for comparison between systems and services.

Compared with other areas in digital library research, as Theng, Mohd-Nasir, and Thimbleby (2000a) pointed out, “Little work is being done to understand the purpose and usability of digital libraries” (p. 238). Borgman et al. (2000) stated, “Relatively little
work has been done on evaluating the usability of digital libraries in any context” (p. 229). The same observations are made by Blandford, Stelmaszewska, and Bryan-Kinns (2001, p.181), Brogan (2003, p. 3), and Blandford, Buchanan, and Jones (2004, p. 69). Thomas (1998) observed that “Although there is a growing body of literature concerning human-computer interaction and information systems, up-to-date empirical research in the area of usability is lean at best” (p. 2). Blandford and Buchanan (2002b) call for further work on methods for analyzing usability, including an understanding of how to balance rigor, appropriateness of techniques, and practical limitations. Popp (2001) also found that the literature on usability testing of library Web sites was small.

I have been motivated by those observations to select this challenging area for my dissertation research.

### 1.2 Problem Statement

This research seeks to develop and evaluate methods and instruments for assessing the usability of digital libraries. Digital libraries offer great potential but, compared to other areas of study, little work is being done to study usability. Usability assessment is one kind of digital library evaluation. Evaluation of digital libraries is still in the formative stage and many concepts remain to be clarified. I hope that this dissertation can contribute to a better understanding of usability and provide a review of the methods that have been applied, while proposing a suite of methods for evaluating usability for academic digital libraries.
1.3 Research Goals

This dissertation study seeks to evaluate methods for assessing the usability of a digital library and to develop a suite of methods for evaluating academic digital libraries. The research goals are:

1. to develop a model of evaluating usability for academic digital libraries;
2. to use the model for the development of measures, instruments, and methods for evaluating academic digital libraries;
3. to test the measures, instruments, and methods by applying them in two academic digital libraries;
4. to generalize the model, instruments, and methods for use in academic libraries;
5. to study the relationships among effectiveness, efficiency, and satisfaction.

1.4 Organization of This Dissertation

This dissertation contains seven chapters:

Chapter 1 is an introduction, containing background, problem statement, and research goals.

Chapter 2 provides a review of the literature, including definitions of digital library, dimensions of usability, techniques of usability evaluation, selected usability tests of digital libraries, and synthesis of findings. These lead to the design of a model proposed in this dissertation.

Chapter 3 presents a usability evaluation model, including the criteria of effectiveness, efficiency, satisfaction, and learnability. Satisfaction is further examined
by ease of use, organization of information, labeling, visual appearance, content, and error correction. Each criterion is defined and its measure is proposed. Figure 2 is an illustration of the model. The study found that an interlocking relationship exists among these criteria. However, each criterion has its own emphasis and should be measured separately.

Chapter 4 describes methods of data collection and analysis, including site selection, stages, subjects, tasks, instruments, setting, procedure, and pilot study. The instruments of this study are presented in the Appendixes.

Chapter 5 presents the results of the study, including both quantitative and qualitative data in both stages.

Chapter 6 is a discussion of the findings.

Chapter 7 concludes the dissertation by presenting research contributions and limitations, and suggests future research that could extend from this study.
Chapter 2

Literature Review

This chapter begins with a review of digital library definitions, which tend to be loose and full of variations. It is followed by a discussion of the dimensions of usability, techniques of usability evaluation, selected usability tests of digital libraries, and a synthesis of findings. These literature reviews lead to the design of an evaluation model and instruments, to be proposed in the next chapter.

2.1 Definitions of Digital Library

There are many different views in the literature as to the actual nature of digital libraries. The National Science Foundation, for instance, supports a wide variety of projects called digital library. Lynch (2005) pointed out that “The field of digital libraries has always been poorly-defined.” The lack of a clear definition of digital library is no longer viewed as a problem, but rather as a feature of the digital library. This dissertation does not intend to provide a comprehensive collection of definitions of the digital library, but rather a number of representative definitions.

The original program announcement of the Digital Library Initiative (DLI or DLI-1) in 1994, sponsored by the NSF, DARPA, and NASA, stated:

Digital libraries basically store materials in electronic format and manipulate large collections of those materials effectively. Research into digital libraries is research into network information systems, concentrating on how to develop the necessary infrastructure to effectively mass-manipulate the information on the Net. The key technical issues are how to search and display desired selections from and across large collections.

Lesk (1997) views digital libraries as “organized collections of digital information” (p. 1). Arms (2000) views them as “managed collection of information,
with associated services, where the information is stored in digital formats and accessible over a network” (p. 2). Witten and his colleagues (Witten, Bainbridge, & Boddie, 2001; Witten & Bainbridge, 2002) state that:

[A digital library is] a focused collection of digital objects, including text, video, and audio, along with methods for access and retrieval, and for selection, organization and maintenance of the collection.

This definition deliberately accords equal weight to end users (access and retrieval) and librarian (selection, organization, and maintenance).

Wolfram and Xie (2002) report that, until recently, digital libraries have focused primarily on well-defined topic collections for specialized audiences, analogous to physical library special collections. They report that general-audience digital libraries are emerging. Examples of those general-audience digital libraries are those growing numbers of state-supported services such as the BadgerLink service in Wisconsin, the Washington State Digital Library Resources, the Inspire service in Indiana, the Alabama Virtual Library, DelAWARE (The Digital Library of the First State), the Tennessee Electronic Library, the Michigan Electronic Library, and the Kentucky Commonwealth Virtual Library (Xie & Wolfram, 2002). Those digital libraries provide access to general and specific interest information sources such as full-text content of magazines, newspapers, professional journals, selected monograph resources, and other published and unpublished information.

Lyman (2001) classifies digital libraries into four kinds: (a) digitized libraries, which include digitized collections of analog documents in all media; (b) born digital libraries, which are a cluster of document types originally created in digital form; (c) data libraries, which include a large variety of different kinds of data collections managed by
databases in servers; and (d) digital communications, which contain a group of electronic
documents such as the Web, e-mail, and Usenet.

Francisco-Revilla, Shipman, Furuta, Karadkar, and Arora (2001) report that
digital libraries are increasingly being defined as ones that collect pointers to WWW-
based resources rather than include the resources themselves. Greenstein (2000) shares
this view and says that the digital library is known less for the extent and nature of the
collections it owns than for the networked information space it defines through its online
services. Paepcke et al. (1996) state that a digital library provides a single point of access
to a wide range of autonomously distributed sources. These views can be seen as calling
include a library’s Web site as one kind of digital library:

Digital libraries include personal, distributed, and centralized collections such as
online public access catalogs and bibliographic databases, distributed document
databases, scholarly and professional discussion lists and electronic journals,
other online databases, forums, and bulletin boards. (p. 672)

As stated by Fox, Suleman, Gaur, and Madalli (2005), “The boundaries between
DLs and other Web-based systems are no longer well defined” (p. 32).

In addition, digital libraries may be seen as new forms of information institutions,
multimedia information retrieval systems, or information systems that support the
creation, use, and searching of digital content (Borgman, Sølvberg, & Kovács, 2002).
Digital libraries also represent a new infrastructure and environment that has been created
by the integration and use of computing, communications, and digital content on a global
scale destined to become an essential part of the information infrastructure in the 21st
century (DELOS, 2004). Chen (2003) shares these views and envisions digital libraries
becoming effective repositories of global knowledge resources, encompassing all aspects
of human activities, from industries to governments, and from education to research. Ioannidis (2005) further points out that the emerging role of digital libraries now calls for a renaming of the field, as its original name is no longer appropriate. Besides libraries, it also encompasses digital archives and museums; it has moved from capturing just text to dealing with multimedia objects with general knowledge, semantics and behavior often embedded in them.

Although there has not been a consensus on the definition of digital library in the research community, the Digital Library Federation [DLF] (1999), representing the practical community, defines digital library as follows:

Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities.

DLF’s definition extends from other definitions that focus on the networked nature or organized digital collections to include a professional staff’s expertise and emphasizes the services that staff can bring to a digital library. A digital library differs from a digital collection or digital archives in that it provides services. Borgman (1999) provides a review of digital library definitions and concludes that, in general, researchers focus on digital libraries as content collected on behalf of user communities, while practicing librarians view digital libraries as institutions or services.

Xie and Wolfram (2002) provide a good analysis of different digital library stakeholders’ perceptions:

To librarians, digital libraries carry out the functions of libraries in a new way; to computer scientists, a digital library is a distributed text-based information system – a collection of distributed information services, a distributed space of interlinked information or a networked multimedia information systems; to end
users, digital libraries are regarded as being similar to the WWW with improvements in performance, organization, functionality, and usability. (p. 1085)

It is therefore summarized in this dissertation that a digital library in the 21st century has the following characteristics:

- It is an organized and managed collection of digital information/objects;
- is accessible over Internet or server;
- is a global information infrastructure; and
- should offer service.

This is not to say that a digital library must be an international collaboration project, but rather to highlight the necessity of the modern digital library’s looking at its mission on a global scale.

In terms of the evaluation of a digital library, as Borgman, Solvberg, and Kovács (2002) state, “Digital libraries are not ends in themselves; rather, they are enabling technologies for digital asset management … electronic publishing, teaching and learning, and other activities. Accordingly, digital libraries need to be evaluated in the context of specific applications” (p. 7). Ioannidis (2005) states, “Evaluation is different if a digital library is viewed as an institution, as an information system, as a new technology or as a combination of new services” (p. 261).

This dissertation selects two academic digital libraries as test sites to evaluate their usability. Academic digital libraries, such as the Rutgers University Libraries Web site (http://www.libraries.rutgers.edu/) and the Queens College Library Web site (http://qcpages.qc.edu/Library/) provide organized and well-managed pointers to their rich local holdings, electronic resources, and services that support their target users’ teaching, learning, and research. Through an academic library’s Web site, users may
access a collection of e-journals and e-books (digital objects). Many of those electronic
collections are made available to the library community via international collaboration.
The e-collections available through the library’s Web sites are accessible over Internet.
Furthermore, such academic libraries provide services to their users. The library Web
sites of Rutgers and Queens are therefore a type of digital library and have been chosen
as test sites. To measure the usability of these two sites, this dissertation will design a
suite of instruments and methods which, it is hoped, will also serve as a model for the
evaluation of other academic digital libraries.

2.2 Dimensions of Usability

Usability is a multidimensional construct that can be examined from various
perspectives. It is also an elusive concept and is determined by the tasks, the users, the
product, and the environment. We have seen in the literature that the term usability has
been used broadly and means different things to different people.

One perspective is to look at usability from an interface effectiveness point of
view. This view makes sense as usability has a theoretical base in human-computer
interaction. Back in 1994, Rubin (1994) pointed out that system designers were slow to
respond to the guidelines established from research in human-computer interaction.
Today, there are many usability studies focusing on interface design. Kim (2002) found
that “the difference between interface effectiveness and usability is not clear” (p. 26).
Interface is one of the most important aspects of usability as it is the medium through
which users communicate and interact with the system.
Usability can also be related to usefulness, usableness, and ease of use. The most concise definition of usability is “fit for use” (American Heritage Dictionary of the English Language, 2000, p. 1894). The Usability Professionals’ Association (2005) defines usability as follow: “Usability is the degree to which something - software, hardware or anything else - is easy to use and a good fit for the people who use it.” Gluck (1997) differentiates between usefulness and usableness. Usableness refers to functions such as “Can I turn it on?” “Can I invoke that function?” Usefulness refers to functions such as “Did it really help me?” “Was it worth the effort?” Blandford and Buchanan (2003) also discuss the difference between usefulness and usableness. Usefulness is generally taken to mean “supporting the required functionality.” Usableness, at its simplest, means “can be used.” Pearrow (2000) argues strongly that “a Web site that is not usable is useless” (p. 4).

Landauer (1995) defines usability as “ease of operation” and usefulness as “serving an intended purpose” although Landauer also comments that the two are hard to separate in the context of evaluation. Davis and his colleagues (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989) also make a distinction between usefulness and ease of use. In their Technology Acceptance Model (TAM) they defined perceived usefulness as the extent to which an information system will enhance a user’s performance and perceived ease of use as the extent to which a person believes a system will be free of effort. Furtado, Furtado, Mattos, and Vanderdonckt (2003) also consider usability from an ease of use point of view and add that usability should also include ease of learning. Grudin (1992), however, considers that usefulness is the issue of whether the system can be used to achieve some desired goal, and breaks it down into utility and usability, where utility is
the question of whether the functionality of the system in principle can do what is needed, and usability is the question of how well users can use that functionality.

Shackel (1991, p. 24) reports that the definition of usability was probably first attempted by Miller (1971) in terms of measures for “ease of use,” and these were first fully discussed and a detailed formal definition was attempted by Shackel (1981, 1984):

the capability in human functional terms to be used easily and effectively by the specified range of users, given specified training and user support, to fulfill the specified range of tasks, within the specified range of environmental scenarios. (Shackel, 1984, p. 53-54)

Perhaps the most widely cited definitions are the ones of ISO and Nielsen. The International Standards Organization (1994) defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” (p. 10). Nielsen (1993) points out that usability has five attributes: (a) learnability: The system should be easy to learn, (b) efficiency: The system should be efficient to use, (c) memorability: The system should be easy to remember so that the casual user is able to return to the system after some period of not having used it, without having to learn everything all over again, (d) errors: The system should have a low error rate, so that users make few errors during their use of the system and can easily recover from any error they may make, and (e) satisfaction: The system should be pleasant to use. Brinck, Gergle, and Wood (2002) share a similar perspective that usability is: (a) functionally correct: The system correctly performs the functions that the user needs; (b) efficient to use; (c) easy to learn; (d) easy to remember; (e) error tolerant: Errors are easily prevented, detected, identified, and corrected; and (f) subjectively pleasing. In addition, Booth (1989) outlines four factors of usability: usefulness, effectiveness (ease of use), learnability, and attitude (likeability).
Shackel (1986) suggests that usability has the criteria of effectiveness, learnability, flexibility, and user attitude. Quesenbery (2001) declares that usability has five E’s: effective, efficient, engaging (i.e., pleasant and satisfying to use), error tolerant, and easy to learn. Hix and Hartson (1993) classify usability into initial performance, long-term performance, learnability, retainability, advanced feature usage, first impression, and long-term user satisfaction. Hix and Hartson are unique in that they take one step further to differentiate performance and satisfaction into initial and long-term measures.

Blandford and Buchanan (2003) break usableness into five components: (a) how efficiently and effectively users can achieve their goals with a system, (b) how easily users can learn to use the system (learnability), (c) how well the system helps the user avoid making errors, or recover from errors, (d) how much users enjoy working with the system or whether they find it frustrating, and (e) how well the system fits within the context in which it is used. Blandford and Buchanan’s definition of usableness is similar to those definitions cited in this paragraph about usability.

It is worth noting that satisfaction is the most frequently cited attribute of usability while usefulness is the attribute often overlooked (Thomas, 1998).

Gould (1988) breaks usability down into more components, including system performance (reliability, responsiveness), system functions, user interface, reading materials, language translation, outreach program, ability for customers to modify and extend, installation, field maintenance and serviceability, advertising, and support group users. Rushinek and Rushinek (1986) found that system responsiveness (response time) is the most important variable affecting users’ happiness.
Karoulis and Pombortsis (2003) suspect that usability (effectiveness, efficiency, and satisfaction) and learnability of educational environment are positively correlated and wonder how far one affects the other, although they did not actually carry out a study to examine this possible correlation.

Guenther (2003) defines usability as ease of navigation, ability with minimal clicks to complete tasks, appealing visually yet easy to understand, and providing the appropriate level of interaction in order to facilitate the completion of tasks while keeping users engaged.

Usability can also be grouped into two large categories: inherent usability (Kurosu & Kashimura, 1995) and apparent usability (Kurosu & Kashimura, 1995; Tractinsky, 1997). Inherent usability is mainly related to the functional or dynamic part of interface usability. It focuses on the way to make the product easy to understand, easy to learn, efficient to use, less erroneous, and pleasurable. On the other hand, apparent usability relates more to the visual impression of the interface. At times, inherent usability and apparent usability may be contradictory (Fu, 1999). For example, in Web page design, graphics enhance apparent usability but slow down the system. Pearrow (2000) argues that form and function are at the opposite ends of a continuum. A truly great Web site and/or digital library should be both aesthetically pleasing and truly usable.

In terms of a usability framework, Thomas (1998) sorts usability attributes into three categories: outcome, process, and task. We may apply Thomas's grouping and categorize usability criteria as follows: the outcome group includes effectiveness, efficiency, and satisfaction; the process group includes ease of use, interface, learnability, memorability, and error recovery; the task group includes functionality and compatibility.
Saracevic (2005) breaks down usability in terms of content, process, format, and overall assessment. Shackel (1991) views usability as the dynamic interplay of four components: user, task, tool, and environment. This relationship may be illustrated as shown in Figure 1 which is a modified version from Shackel (1991, p.23), Bennett (1972, 1979) and Eason (1981). The emphasis of this modified figure is on the interplay relationships among user, task, and tool. All are in the context of the environment.

**Figure 1. The Four Principle Components in a Human-Machine System**

Usability has user focus. Dumas and Redish (1993) point out that with good usability “people who use the product can do so quickly and easily to accomplish their task” (p. 4). Guillemette (1995) refers to usability as “the degree to which an information system can be effectively used by target users in the performance of tasks” (p. 215). Fowler (1997) and Clairmont, Dickstein, and Mills (1999) state, “Usability is the degree to which a user can successfully learn and use a product to achieve a goal.”

Usability is different from functionality. Dumas and Redish (1993) use the videocassette recorder (VCR) as an example to illustrate the difference between the two: VCRs may have high functionality (the feature works as it was designed to work) but
they have low usability (people cannot use them quickly and easily to accomplish their task).

Usability is not equivalent to accessibility. Accessibility involves making Web site content available to and usable by people with disabilities.

Table 1 compares various perspectives on the attributes of usability.

Usability has several aspects, including interface design, functional design, data and metadata, and computer systems and networks (Arms, 2000). Usability is a property of the total system. All the components must work together smoothly to create an effective and easy-to-use digital library.

Usability can be tackled from various directions. Blandford and Buchanan (2002a) suggest that usability is technical, cognitive, social, and design-oriented and it is important to bring these different perspectives together, to share views, experiences and insights. Spinuzzi (1999) states that usability is not an isolated quality but a feature diffused throughout the ecology of technology. Digital library development involves interplay among people, organization, and technology. The usability issue involves the system as a whole.

In order to design systems for a variety of users, Lamb (1995) suggests that usability issues should be extended beyond interface usability to include content usability, organizational usability, and inter-organizational usability.
### Table 1

**Attributes of Usability**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booth (1989)</td>
<td>usefulness, effectiveness, learnability, attitude</td>
</tr>
<tr>
<td>Brinck et al. (2002)</td>
<td>functionally correct, efficient to use, easy to learn, easy to remember, error tolerant, and subjectively pleasing</td>
</tr>
<tr>
<td>Clairmont et al. (1999)</td>
<td>successfully learn and use a product to achieve a goal</td>
</tr>
<tr>
<td>Dumas &amp; Redish (1993)</td>
<td>perform tasks quickly and easily</td>
</tr>
<tr>
<td>Fowler (1997)</td>
<td>successfully learn and use a product to achieve a goal</td>
</tr>
<tr>
<td>Furtado et al. (2003)</td>
<td>ease of use and learning</td>
</tr>
<tr>
<td>Gluck (1997)</td>
<td>useableness, usefulness</td>
</tr>
<tr>
<td>Guenther (2003)</td>
<td>ease of navigation, efficiency, visual appeal, easy to understand, appropriate level of interaction</td>
</tr>
<tr>
<td>Guillemette (1995)</td>
<td>effectively used by target users to perform tasks</td>
</tr>
<tr>
<td>Hix &amp; Hartson (1993)</td>
<td>initial performance, long-term performance, learnability, retainability, advanced feature usage, first impression, and long-term user satisfaction</td>
</tr>
<tr>
<td>ISO (1994)</td>
<td>effectiveness, efficiency, satisfaction</td>
</tr>
<tr>
<td>Kengeri et al. (1999)</td>
<td>effectiveness, likeability, learnability, usefulness</td>
</tr>
<tr>
<td>Kim (2002)</td>
<td>interface effectiveness</td>
</tr>
<tr>
<td>Nielsen (1993)</td>
<td>learnability, efficiency, memorability, errors, satisfaction</td>
</tr>
<tr>
<td>Oulanov &amp; Pajarillo (2002)</td>
<td>affect, efficiency, control, helpfulness, adaptability</td>
</tr>
<tr>
<td>Quesenbery (2001)</td>
<td>effective, efficient, engaging, error tolerant, easy to learn</td>
</tr>
<tr>
<td>Shackel (1981)</td>
<td>ease of use, effectiveness</td>
</tr>
<tr>
<td>Shackel (1986, 1991)</td>
<td>effectiveness, learnability, flexibility, user attitude</td>
</tr>
</tbody>
</table>
In addition to those aspects, usability can be examined from the perspectives of graphic design, navigation, and content (Spool, Scanlon, Snyder, Schroeder, & DeAngelo, 1999). Turner (2002) sorts usability into navigation, page design, content, accessibility, media use, interactivity, and consistency. Clausen (1999) declares that a high quality Web page should display accuracy, authority, currency, relevant links, browsability, navigation, consistency, use of frames, use of graphics, connectivity, interactivity, user-friendliness, and originality.

There are several approaches in evaluating usability. Eberts (1985) sorts them into four categories: empirical, anthropomorphic, cognitive, and predictive modeling approaches. The empirical approach emphasizes obtaining a representative sample of target users and analyzes user performance and attitudes under specified conditions. The anthropomorphic approach evaluates user-system interaction in comparison to effective interpersonal communication. The cognitive approach applies the theory and methodology of cognitive psychology to interface design. The predictive approach uses an analytic tool which examines or manipulates an abstract representation of the interface in order to forecast usability. Blandford, Keith, Connell, and Edwards (2004), on the other hand, roughly group usability evaluation into two kinds: empirical and analytical. Empirical techniques involve testing systems with users, whereas analytical techniques involve the assessment of systems by personnel using established theories and methods. In addition, Levi and Conrad (2002) divide usability testing into three styles: exploratory testing, threshold testing, and comparison testing. Exploratory testing examines a system and looks for areas of user confusion, slow-down, or mistakes. Such testing is performed with no particular preconceived notions about where the problems lie or what form they
may take. Exploratory testing can be used at any point in the development life cycle, but is most effective when implemented early and often. Threshold testing measures the performance characteristics of a system against predetermined goals. Threshold testing typically accompanies a beta release. Comparison testing measures the usability characteristics of two approaches or designs to determine which better suits users’ needs. Comparison testing is usually done at the early prototyping stage.

In conclusion, usability can be defined as (Pearrow, 2000):

Usability is the broad discipline of applying sound scientific observation, measurement, and design principles to the creation and maintenance of Web sites in order to bring about the greatest ease of use, ease of learnability, amount of usefulness, and least amount of discomfort for the humans who have to use the system. (p. 12)

Although Pearrow’s definition is about Web sites, the definition is applicable to digital library and information systems as well. Pearrow’s definition covers the criteria of ease of use, learnability, usefulness, and comfort. The definition also emphasizes the application of scientific observation and measurement methods to assess usability.

Based on the discussion of the dimensions of usability in this chapter, a model to evaluate usability is proposed in Chapter 3. The criteria in the proposed model include quantifying elements, such as effectiveness, efficiency, and learnability, as well as Likert scales and qualitative data on satisfaction. The satisfaction criterion is further examined by ease of use, organization of information, labeling and terminology, visual appearance, content, and error correction.
2.3 Techniques of Usability Evaluation

There are a number of methods to evaluate usability. The techniques include formal usability testing, usability inspection, card sort, category membership expectation, focus groups, questionnaires, think aloud, analysis of site usage logs, cognitive walkthrough, heuristic evaluation, claims analysis, concept-based analysis of surface and structural misfits (CASSM), paper prototyping, and field study (Askin, 1998; Blandford, Keith, Connell, & Edwards, 2004; Campbell, 2001; Kantner & Rosenbaum, 1997; Keith, Blandford, Fields, & Theng, 2003; Levi & Conrad, 2002; Nielsen & Mack, 1994; Pearrow, 2000; Popp, 2001; Rosson & Carroll, 2002; Snyder, 2003).

Card sort is one of the simplest yet most useful. It requires few resources, needs little time to complete, and provides quite a bit of insight. It should be used when determining where particular pieces of information should go on the site. Card sort is best used when participants have never seen the site or when the site is undergoing a major redesign. The objective of card sort is to see if the site architecture makes sense to the user. It is more commonly used as a design tool for building menu trees.

Formal usability testing is the most productive technique in that the information gained is the most detailed and closest to the actual user. Formal usability testing has five characteristics: (a) The primary goal is to improve the usability of a product; (b) The participants represent real users; (c) The participants do real tasks; (d) The things participants do and say are observed and recorded; and (e) The data are analyzed, the real problems are diagnosed, and recommendations to fix those problems are made (Dumas & Redish, 1993).
Usability inspection applies to early designs, well-developed designs, and deployed systems; is usually less expensive than lab-based formal usability testing; does not employ real users; is expert-based (conducted by trained usability engineering practitioners); is often, but not always, guided by user tasks; and has the goal of predicting usability problems that users will encounter in real usage (Hartson, Shivakumar, & Pérez-Quinones, 2004).

Paper prototyping is a variation of usability testing where representative users perform realistic tasks by interacting with a paper version of the interface that is manipulated by a person “playing computer,” with a separate page for every choice on the screen (Snyder, 2003).

Heuristic evaluation is also sometimes called a *usability audit* or *heuristic expert review*. A small group of evaluators analyzes the interface of the site by a list of recognized usability principles, called *heuristics* (or rules of thumb). It is probably the most widely used usability evaluation technique (Blandford, Keith, Connell, & Edwards, 2004, p. 28) because it is perceived to yield reasonable benefits for low cost. As Nielsen and Mack (1994) state, it is easy, fast, and cheap. However, it tends to focus on micro features of an interface rather than the global picture. Heuristic evaluation is often best done at the beginning of a project and then repeated at later stages of development (Tom, 2000). Nielsen (n.d.) found that three to five evaluators detected most of the usability problems.

Cognitive walkthrough is another method that involves expert evaluators. The evaluators design specific task scenarios. The user’s goals and purpose for each task are defined and tasks are broken down into relatively small pieces. The evaluator’s role is to...
play the part of the user working with the site, noting problems, paths, and barriers, essentially reviewing the ease of learning of the site (Wharton, Rieman, Lewis, & Polson, 1994). Cognitive walkthrough is a relatively cheap method of cleaning up a Web site or a digital library. Since it does not involve actual users, it can be conducted any time and as frequently as desired. Cognitive walkthrough is more limited in scope than heuristic evaluation because of its rigid structure but, at the same time, it provides a clear structure for conducting the analysis once user profiles and tasks have been defined. The greatest challenge of cognitive walkthrough is in specifying tasks. Both heuristic evaluation and cognitive walkthrough evaluation address surface features of usability well, but do not identify deeper issues such as how users formulate good queries, evaluate results, and how users interact with the information.

Claims analysis is less structured than cognitive walkthrough. It is more difficult to learn than heuristic evaluation and cognitive walkthrough (Blandford, Keith, Connell, & Edwards, 2004, p. 32), but it supports the analyst in thinking about usability issues more deeply. Claims analysis provokes thinking about why things are the way they are, and how they could be different. In this method, the usability engineer identifies significant features in a design and generates hypotheses about the consequences these features might have for users engaged in activities.

CASSM (concept-based analysis of surface and structural misfits) considers design in terms of concepts: the concepts the user is working with, those implemented within the system, and those represented at the interface. The analysis focuses on the quality of fit between the user and system concepts (Blandford, Keith, Connell, & Edwards, 2004; Connell, Green, & Blandford, 2003). CASSM does not deal with
usability issues at the levels of detail of heuristic evaluation, cognitive walkthrough, or claims analysis. It takes more of a broad-brush approach in considering key user and system concepts. It is also more difficult to learn than heuristic evaluation or cognitive walkthrough. While both claims analysis and CASSM probe deeper issues, claims analysis is more demanding of the analyst and delivers a wider range of insights.

Field study has applicability to Web usability studies. A field study is a semi-structured period of observation of users in their natural environment. This technique is used in the beginning of the design process to gather understanding of current limitations and areas for improvement.

Thinking aloud is actually subsumed by other tools and techniques. It helps to capture a user’s moments of confusion or hesitation and to notice user preconceptions and errors.

The questionnaire is powerful as it allows the investigator to gather a great deal of information from a large number of participants very cheaply.

The objective of interview is to gather information about user’s expectations, needs, thoughts, past experience, and other information related to the site in a semi-formal format.

Focus groups are an inquiry method that looks at users’ opinions or perceptions. In a usability study, focus groups are used in several different ways. Some conduct focus groups to look at a site initially (usually copies rather than live sites) and discuss issues about the site. This preliminary information is used to create prototypes for the site. Some use focus groups immediately after conducting a formal usability test (Palmquist, 2001; Thomsett-Scott, 2004). The weakness of focus groups in usability testing is that
participants in focus groups reveal what they think they think and what they think they do, but not what they actually believe or what they actually do. The necessity of watching what actual users are doing cannot be replaced by focus groups.

Log file analysis is a time-consuming activity. There are vast amounts of data generated even in a short evaluation period. The usage data is data only, and the reasoning behind the usage numbers could only be speculation, unless the log file analysis is used with another usability evaluation method.

According to Popp's survey (2001) of members of the Association of Research Libraries, cognitive walkthrough is the most frequently used method (52 %), followed by individual staff review (50 %), focus groups (44 %), online survey (40%), usability testing (37 %), log analysis (37%), and others. Most libraries used a combination of more than one method.

2.4 Selected Usability Tests of Digital Libraries

A review of the literature indicates that the areas of usability testing for digital libraries have covered breadth of coverage, content, navigation, functionality, utility, interface, metadata appropriateness, awareness of library resources, usefulness, accessibility, and satisfaction. The methods employed include formal usability testing, heuristic evaluation, usability inspection, questionnaires, focus groups, surveys, interviews, transaction log analysis, walkthrough, card sorting, and claims analysis.

The National Taiwan University Library used questionnaires to survey 1784 users on usability (Lan, 2001). They found the site's usability problems are mainly in the areas of information architecture and in the browsing and searching mechanism. The study of
CUNY+ (Oulanov & Pajarillo, 2002) also employed a questionnaire as the primary method of usability assessment. The authors conducted a two-phase study to compare the usability of text-based and Web-based CUNY Web sites. The criteria used were affect, efficiency, control, helpfulness, and adaptability. Ten students were invited to participate using a random sampling method.

Adams and Blandford (2002) reported on their study of accessibility at a large London-based hospital. They conducted focus groups and in-depth interviews with 73 hospital clinicians. Fifty percent of the participants were nurses, while the other 50% were senior and junior doctors, consultants, surgeons, managers, and IT department members. The study focused on two themes: the perceived effectiveness of traditional and digital libraries as clinical resources, and the impact of clinician status on control over and access to information. The participant responses indicated that digital library technology provides remote access to materials, but the system’s usability is poor and it is time-consuming to access information.


Sumner and Dawe (2001) studied the usability of the Digital Library for Earth System Education (DLESE, www.DLESE.org) focusing on its role in the process of educational resource reuse. One finding was that the design of the search results page is critical for supporting resource comprehension. Also, the library’s metadata plays a
central role in documenting the resource enough to support comprehension and modification processes.

Sumner, Khoo, Recker, and Marlino (2003) used DLESE and National Science Digital Library (NSDL, www.NSDL.org) to study usability issues. The purpose of this study was to identify educators' expectations and requirements for the design of educational digital collections for classroom use. A series of five focus groups was conducted with a total of thirty-six teachers and two librarians to review eighteen Web sites. The participants indicated that content quality, advertising, bias, and design were important factors influencing their perceptions.

Hartson, Shivakumar, and Pérez-Quiñones (2004) applied the usability inspection method to evaluate the design and functionality of the Networked Computer Science Technical Reference Library (NCSTRL, http://www.ncstrol.org). They found NCSTRL’s design was apparently functionally oriented rather than an approach based on user task threads. Another finding of the usability inspection was about the terminology used in NCSTRL. There was jargon and the use of terms was designer-centered rather than user-centered.

The evaluation of DeLiVer applied a mixture of methods, including transaction log analysis, surveys, interviews, focus groups, and formal usability testing to measure accessibility (Bishop, 2001; Neumann & Bishop, 1998). The evaluators learned that triangulation of data is crucial. The evaluation process has allowed them to pursue the different social issues surrounding digital library use and well as to deal with specific usability issues.
The evaluation of eLibraryHub also applied the approach of triangulation of data (Theng, Chan, Khoo, & Buddharaju, 2005). The study employed both quantitative and qualitative methods, including questionnaires and claims analysis, to study effectiveness, usefulness, and satisfaction. Recommendations were made for further improvement to eLibraryHub.

The University of Arizona Library applied a number of methods to evaluate the usability of the Library Web site, SABIO, including heuristic evaluation, walk-through, card sorting, and formal usability testing (Dickstein & Mills, 2000; Veldof, Prasse, & Mills, 1999). Heuristic evaluation was used to systematically inspect user interface; walk-through was used to explore and to envision users’ problems in the prototype stage; card sorting was used to assess organization and menu structure; and formal usability testing was used to observe the real users’ use of the site.

The study at Nanyang Technological University in Singapore was conducted in two phases (Peng, Ramaiah, & Foo, 2004). In the first phase a questionnaire was distributed to 100 students; eighty-eight responses were received. The questionnaire contained questions pertaining to the ten heuristic principles of Nielsen (1994). The second phase was formal usability testing by 52 students who also participated in the first phase. The criteria of usability testing in the second phase were efficiency, learnability, ease of use, memorability, error rates per session, nature of errors, system’s aesthetics, navigability, and use of technical jargon. It found that undergraduates spent more time to complete the given tasks than did postgraduates.

Dorward, Reinke, and Recker (2002) evaluated Instructional Architect, which aims to increase the utility of NSDL resources for classroom teachers. The methods they
employed included formal usability testing and focus groups. The evaluation centered on interface design and contents. It was suggested that an introductory tutorial, better graphics, and a preview screen should be incorporated. The usability study of the Belgian-American Research Collection at the University of Wisconsin Digital Collections Center also used focus group and formal usability testing (Clark, 2004). The focus group consisted of seven graduate students and the formal usability testing had five users (students, staff, and faculty). There was also an exit survey about demographics, satisfaction, and general feelings about the system. The study results led the system designer to present information more logically and intuitively for the users.

The University of the Pacific applied the formal usability testing technique to measure students’ awareness of library resources (Krueger, Ray, & Knight, 2004). They recruited 134 students to perform eight tasks, including locating an article, locating a journal, finding call number of a book, finding overdue information, finding a biography, and connecting from home. They found 45% of the participants were familiar enough with library resources and 34% were regular users of library Web resources. They also found that the majority of their students know how to search for books in their OPAC but many flounder when asked to find similar information for journals. Another lesson the librarians learned was that they should have employed a smaller number of samples using purposeful sampling. This would have allowed them to gather more useful data by targeting small groups of students representing demographic characteristics of interest.

Florida International University also applied the formal usability testing technique to study its library’s Web site (Hammill, 2003). The goal was to determine whether the design and organization of the top page of the site allow users to locate information easily.
The study focused on three areas: catalog searching, article searching, and library services. The study recruited 52 participants, including 26 undergraduates, 14 graduates, 5 faculty, and 7 others. The study collected performance data such as number of clicks to find the answer. The comments from the think-aloud process and the post-test interview and questionnaire helped to evaluate the users' satisfaction. As a result of the study, the librarians found that solely testing the top page limited their abilities to measure navigation clearly. They also found that they should have focused on a smaller number of subjects and gone deeper into issues instead of having many users for a short period of time.

Minnesota’s Foundations Project is a multi-agency collaboration to improve access to environmental and natural resources information (Quam, 2001). It conducted two usability studies: one to study ease of use of the interface and one to study the usefulness of controlled vocabulary. A formal usability test was the primary method used, supplemented by pretest and post-test questionnaires. There were a total of six participants. The researchers found that controlled vocabulary is important in facilitating access. The feedback from the usability test resulted in changes in the interface to lessen confusion.

In addition to the University of the Pacific, Florida International University, and Minnesota’s Foundation Project, a sizable number of university libraries and other academic digital libraries have applied a formal usability testing method as their primary usability evaluation method. These include Arizona State University West (Collins & Aguiñaga, 2001), Auburn University (McDonald, 2001), Dalhousie University (Gullikson et al., 1999), Eastern Illinois University (Gibbs, 2001), Memorial University of
Newfoundland (McGillis & Toms, 2001), MIT Libraries (Hennig, 1999), National Library of Medicine (Marill, 2001), Roger Williams University (McMullen, 2001a; McMullen, 2001b), University of Arizona (Dickstein, Loomis, & Veldof, 1999; Dickstein & Mills, 2000; Veldof, Prasse, & Mills, 1999), University of Buffalo (Battleson, Booth, & Weintrop, 2001), University of Illinois at Chicago (Augustine & Greene, 2002), University of Minnesota (Dickstein, Loomis, & Veldof, 1999), University of Saskatchewan (Duncan & Fichter, 2004), University of South Florida (Allen, 2002), University of Southern California (Benjes & Brown, 2001), University of Wisconsin (Dickstein, Loomis, & Veldof, 1999), Washington State University (Chisman, Diller, & Walbridge, 1999; Walbridge, 2000), and Western Michigan University (Cockrell & Jayne, 2002).

The Getty Research Institute applied formal usability testing and log analysis to study navigation of its library online catalog (Hessel & Burton-West, 2003). Its subjects included four current library users and four potential users. A program was written to further analyze the search logs provided by the Voyager system on the usage frequency of each index. The study helped the library to redesign default search and the help screen of the catalog.

The University of North Texas Libraries triangulated its usability assessment by using formal usability testing and focus groups (Thomsett-Scott, 2004). The formal usability testing was conducted first and the results formed the basis of the discussion in the focus groups. There were two studies: one was to study the Ask a Librarian sub-site and the other one the Libraries’ home page. The formal usability testing had 30 subjects each. The focus groups had four to six participants each.
The University of Nevada, Las Vegas, employed five different assessment methods: log file analysis, internal survey, heuristic evaluation, formal usability testing, and card sort (Felker & Chung, 2005). This method of triangulation confirmed the findings, even though only seven responses to the survey were received and only ten students participated in the formal usability test. The problems of their Web site centered in three main areas: structure, page setup and layout, and help and documentation.

In addition to the above-mentioned usability evaluations of actual academic digital libraries, the Winona State University Library conducted two focus groups to ascertain the usability of various Web page design elements (Leighton, Jackson, Sullivan, & Dennison, 2003). The first focus group consisted of two students and one faculty member. The second focus group consisted of seven students. The research team designed a series of Web sites for testing the following design features: (a) whether framed pages were more effective than non-framed pages for site navigation, and if so, which location for the frames was most effective; (b) whether links were more effective in narrative text, in lists without contextual information, or in lists with contextual information provided in link annotations; (c) whether length, hierarchical page depth or a same-page compression mechanism (such as a drop-down menu) was more useful, and (d) whether a site specific search engine, an alphabetical site index, or a subject hierarchy was more useful. The research found that frames can be useful but that small frames were preferable, mainly because pages which had large frames often required additional scrolling, due to the space used by the frame itself. There seems to be no difference in success with or without frames. Users in both groups unanimously confirmed that embedded links were difficult to read. Participants of both focus groups also preferred
either the sequence of hierarchical menus or the drop down menus to scrolling the long list of links. Subjects declared that they preferred the hierarchy to the site index.

The usability test at Long Island University also examined the use of frames in the library Web site (Callicott, 2002). The test showed that the framed environment was difficult for some of the more novice Web users to grasp. It also discovered that most users used the "back" button as their prime navigational tool and completely ignored the more precise framed navigational tools. The results of the usability test made clear that the next version of the library Web site will be frames-free. The usability test at the Long Island University library employed a formal usability test technique. The subjects were 5 student workers.

Considering the number of digital libraries or academic library Web sites today, the literature on usability tests for those systems is still small. However, Popp's study (2001) found that interest and involvement in testing library Web sites grew exponentially in the 1990s, with an explosion of work done in 1999. Her survey indicates that there were only two tests in 1996, but there were 15 in 1997, 36 in 1998, and 82 in 1999. The literature review for the present study found continued interest since 1999. Many usability studies reported in library literature were published after 2000. It is clear that libraries are rapidly moving forward in evaluating the usability of their Web sites and are beginning to view usability evaluation as an integral part of Web site development. However, there are still very few usability tests for digital libraries that are not library Web sites.
Table 2 is a review of usability tests in academic digital libraries, summarizing the methods applied, numbers and kinds of subjects, the focal area for the usability test, and the criteria used.

Formal usability testing is the most frequently used method in the academic library setting while for other types of digital library, questionnaire, focus group, and formal usability testing are three most commonly used methods. Many libraries made use of more than one method. Satisfaction is the factor most frequently studied, followed by efficiency, ease of use, effectiveness, and usefulness.
### Table 2

**Usability Assessment at Academic Sites**

<table>
<thead>
<tr>
<th>Site</th>
<th>Methods</th>
<th>Subjects</th>
<th>Areas</th>
<th>Criteria</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM, IEEE-CS, NCSTRL, NDLTD</td>
<td>formal usability test, questionnaire</td>
<td>48 students (38 graduate students, 10 undergraduates)</td>
<td>interface</td>
<td>ease of use</td>
<td>Kengeri et al. (1999)</td>
</tr>
<tr>
<td>ACMDL, NCSTRL, NZDL</td>
<td>questionnaire, heuristic evaluation</td>
<td>45 undergraduates</td>
<td>design and structure</td>
<td>satisfaction</td>
<td>Theng et al. (2000a, 2000b)</td>
</tr>
<tr>
<td>Alexandria</td>
<td>questionnaire, formal usability test</td>
<td>23 students</td>
<td>interface</td>
<td></td>
<td>Thomas (1998)</td>
</tr>
<tr>
<td>Belgian-American Research Collection</td>
<td>focus group, formal usability test, questionnaire</td>
<td>7 graduate students in focus group, 5 users in usability test</td>
<td>design, navigation</td>
<td>satisfaction</td>
<td>Clark (2004)</td>
</tr>
<tr>
<td>Brazilian Digital Library of Computing</td>
<td>questionnaire</td>
<td>21 subjects</td>
<td>self-archiving service</td>
<td>ease of use</td>
<td>Silva et al. (2005)</td>
</tr>
<tr>
<td>California State U. San Marcos Library</td>
<td>observation, focus groups, questionnaire</td>
<td>5 students</td>
<td>navigation</td>
<td></td>
<td>Thompson (2003)</td>
</tr>
<tr>
<td>CUNY+</td>
<td>questionnaire</td>
<td>10 students</td>
<td>interface</td>
<td></td>
<td>Oulanov &amp; Pajarillo (2002)</td>
</tr>
<tr>
<td>DeLiVer</td>
<td>transaction log, survey, interview, focus groups, formal usability test</td>
<td>1900 graduate students, 420 faculty</td>
<td>accessibility</td>
<td></td>
<td>Neumann &amp; Bishop (1998), Bishop (2001)</td>
</tr>
<tr>
<td>Institution</td>
<td>Methods</td>
<td>Participants</td>
<td>Evaluation Items</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------------------------------------</td>
<td>--------------</td>
<td>------------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>DLESE, NSDL</td>
<td>focus groups</td>
<td>36 teachers 2 librarians</td>
<td>design</td>
<td>Sumner et al. (2003)</td>
<td></td>
</tr>
<tr>
<td>eLibraryHub</td>
<td>1) questionnaire 2) claims analysis</td>
<td>107 students and staff 6 subjects</td>
<td>navigation structure information layout</td>
<td>Theng et al. (2005)</td>
<td></td>
</tr>
<tr>
<td>Florida International U.</td>
<td>formal usability test questionnaire</td>
<td>52 subjects</td>
<td>design, organization navigation, vocabulary</td>
<td>Hammill (2003)</td>
<td></td>
</tr>
<tr>
<td>Getty Research Institute</td>
<td>formal usability test log analysis</td>
<td>4 current users 4 potential users</td>
<td>navigation</td>
<td>Hessel, Burton-West(2003)</td>
<td></td>
</tr>
<tr>
<td>Instructional Architect</td>
<td>formal usability test focus group</td>
<td>26 teachers</td>
<td>interface, content</td>
<td>Dorward et al. (2002)</td>
<td></td>
</tr>
<tr>
<td>Living Health Database</td>
<td>questionnaires interviews observations</td>
<td>9 older users 4 deaf users</td>
<td>accessibility</td>
<td>Thompson et al. (2002)</td>
<td></td>
</tr>
<tr>
<td>London Hospital</td>
<td>focus groups interviews</td>
<td>73 clinicians</td>
<td>accessibility</td>
<td>Adams &amp; Blandford (2002)</td>
<td></td>
</tr>
<tr>
<td>Long Island U.</td>
<td>formal usability test</td>
<td>5 student workers</td>
<td>organization navigation</td>
<td>Callicott (2002)</td>
<td></td>
</tr>
<tr>
<td>MARIAN (Virginia Tech)</td>
<td>formal usability test log analysis questionnaire</td>
<td>students, faculty, staff</td>
<td>interface</td>
<td>France et al. (1999)</td>
<td></td>
</tr>
<tr>
<td>McMaster U.</td>
<td>formal usability test</td>
<td>8 students</td>
<td>interface, usefulness ease of use</td>
<td>Detlor et al. (2003)</td>
<td></td>
</tr>
<tr>
<td>Minnesota’s Foundation Project</td>
<td>formal usability test questionnaire</td>
<td>6 subjects</td>
<td>interface, ease of use</td>
<td>Quam (2001)</td>
<td></td>
</tr>
<tr>
<td>MIT</td>
<td>formal usability test</td>
<td>29 (faculty, graduate, undergraduate, staff)</td>
<td>site design</td>
<td>Hennig (1999)</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Methodologies</td>
<td>Participants</td>
<td>Measured Characteristics</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>-------------------------------------------------------</td>
<td>-----------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| Nanyang Technological U.          | 1) questionnaire  
2) formal usability test | 1) 88 students  
2) 52 students | interface, efficiency, learnability, ease of use, memorability | Peng et al. (2004) |
| National Electronic Library for Communicable Diseases | questionnaire  
interview | 30 general public | site structure, organization, terminology, satisfaction | Williams et al. (2004) |
| National Taiwan U.                | questionnaire  
interview | 1784 faculty  
and students | information architecture, satisfaction, browsing & searching mechanism, layout and display | Lan (2001) |
| NCSTRL                            | usability inspection  
focus groups  
questionnaire  
heuristic evaluation  
design walk-through  
card sorting | 3 usability experts  
32 staff | design, interface, functionality, efficiency, satisfaction | Hartson et al. (2004) |
| SABIO                              | formal usability test  
heuristic evaluation  
design walk-through  
card sorting | students | design, efficiency, satisfaction | Dickstein & Mills (2000), Veldof et al. (1999) |
| South London and Maudsley NHS Trust Library | focus groups  
questionnaire  
heuristic evaluation  
observation  
card sorting | 12 students  
32 staff | scope and content, organization, navigation, terminology | Ebenezer (2003) |
| U. of Illinois at Chicago          | formal usability test  
internal survey  
heuristics evaluation  
formal usability test  
card sort | 12 students  
7 survey response  
| U. of Nevada, Las Vegas           | log file analysis  
internal survey  
heuristics evaluation  
formal usability test  
card sort | 7 survey response  
10 students in usability test | design, efficiency, satisfaction | Felker & Chung (2005) |
| U. of North Texas                 | 1) formal usability test  
2) focus groups | 1) 30 students  
2) 4-6 participants | design, effectiveness, efficiency | Thomsett-Scott (2004) |
<p>| U. of Saskatchewan                | formal usability test | 5 faculty and students | label, link, satisfaction | Duncan &amp; Fichter (2004) |</p>
<table>
<thead>
<tr>
<th>Institution</th>
<th>Methodology</th>
<th>Participants</th>
<th>Focus Areas</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>U. of South Florida</td>
<td>formal usability test</td>
<td>26 undergraduates</td>
<td>interface</td>
<td>Allen (2002)</td>
</tr>
<tr>
<td>U. of Southern California</td>
<td>formal usability test</td>
<td>1) 5 faculty, 1 staff, 1 student 2) 2 faculty, 4 staff</td>
<td>architecture, terminology color, navigation</td>
<td>Benjes &amp; Brown (2001)</td>
</tr>
<tr>
<td>U. of the Pacific</td>
<td>formal usability test</td>
<td>134 students</td>
<td>awareness of library resources</td>
<td>Krueger et al. (2004)</td>
</tr>
<tr>
<td>Washington State U.</td>
<td>formal usability test</td>
<td>12 students</td>
<td>navigation</td>
<td>Chisman et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>questionnaire</td>
<td></td>
<td></td>
<td>Walbridge (2000)</td>
</tr>
<tr>
<td>Winona State U.</td>
<td>focus groups</td>
<td>1) 2 students, 1 faculty 2) 7 students</td>
<td>page layout site design</td>
<td>Leighton et al. (2003)</td>
</tr>
</tbody>
</table>
2.5 Synthesis of Results of Digital Library Usability Tests

This section summarizes the principal results of usability tests of academic digital libraries.

Several usability studies on digital libraries report a user “lostness” issue. Theng, Mohd-Nasir, and Thimbleby (2000a) studied the usability of three mature topic-based digital libraries: the ACM Digital Library (http://www.acm.org), the Networked Computer Science Technical Reference Library (http://www.ncstrl.org), and the New Zealand Digital Library (http://www.nzdl.org). They found that 73% of the subjects experienced different degrees of lostness, which means that the digital libraries did not provide appropriate information to help users understand the structure of the digital libraries. Spool, Scanlon, Snyder, Schroeder, and DeAngelo (1999) and Gullikson et al. (1999) also report this lostness problem in their studies. The Alexandria Digital Library (ADL) found that “lost in the library” was a common problem and took steps to resolve this disorientation issue (Buttenfield, 1999). Kengeri, Seals, Harley, Reddy, and Fox (1999) found that their subjects were often confused about the structure of the digital collection, and called for a clear overview of the digital library layout. The clear and logical organization of information in a digital library is indeed an important matter. The present study will examine this issue to determine whether users of the Rutgers University Libraries and Queens College Web sites also experience this problem.

Blandford, Stelmaszewska, and Bryan-Kinns (2001) report navigation disorientation as a result of their usability testing. They found that users could accidentally leave the digital library, following the links to other Web-based resources, and were unaware that they were no longer using the digital library. Indeed, navigation
disorientation is among the biggest frustrations for Web users (Brinck, Gergle, & Wood, 2002). This situation is common, particularly with the increasing provision of digital library portals that provide links to various libraries from one Web site.

The lack of benchmarks for usability testing is a problem. Theng, Mohd-Nasir, and Thimbleby (2000b) reported that they had to make the assumption that if an area scores 75% or above for accuracy it implies that the area is well implemented. The usability testing at MIT Libraries demonstrated that subjects had a 75% success rate (Hennig, 1999). But, they wondered, is 75% high or low? The study at the University of the Pacific (Krueger, Ray, & Knight, 2004) provides effectiveness benchmarks for locating the call number of a book (91%), locating a journal (50%), setting up remote access from home (58%), and locating overdue information (28%). The study of the University of Illinois at Chicago (Augustine & Greene, 2002) contributed both effectiveness and efficiency benchmarks for locating a journal (58%, 1 min. 48 sec., 10.7 clicks), finding an electronic database (75%, 1 min. 18 sec., 5.75 clicks), and finding information. The present study will explore the issues of effectiveness, efficiency, satisfaction, and learnability of a system when locating a book, locating a journal, using electronic databases, locating an e-book, locating an encyclopedia article, and locating information. Thus more benchmarks will be contributed to the literature.

Several usability studies in library Web sites found that libraries’ Web sites are designed from a librarian’s perspective instead of a user’s perspective. McMullen (2001b) discussed this from the interface aspect while Collins and Aguiñaga (2001) examined it from the aspect of the organization of information. Collins and Aguiñaga report that it is helpful to learn that users think differently than librarians about the organization of
information. The feedback received from the usability testing helped the Arizona State University Library to improve the user-friendliness of its Web site. Dickstein and Mills (2000) report that the library is transforming its service model to a more user-focused model. The study at the University of Illinois at Chicago (Augustine & Greene, 2002) found that library users' basic lack of understanding and awareness of library resources impacted their ability more than the organization of the site did.

A number of usability studies report the problem of terminology (Allen, 2002; Augustine & Greene, 2002; Battleson, Booth, & Weintrop, 2001; Bobay, Dallis, Pershing, & Popp, 2001; Callicott, 2002; Chisman, Diller, & Walbridge, 1999; Church, Brown, & VanderPol, 2001; Clark, 2004; Cockrell & Jayne, 2002; Crowley, Leffel, Ramirez, Hart, & Armstrong, 2002; Detlor et al. 2003; Dickstein, Loomis, & Veldof, 1999; Dickstein & Mills, 2000; Gullikson et al., 1999; Hartson, Shivakumar, Pérez-Quiñones, 2004; Hessel & Burton-West, 2003; King, 2001; McDonald, 2001; McGillis & Toms, 2001; McMullen, 2001b; Spivey, 2000; Thompson, 2003; Wegener, Moi, & Li, 2004). The degree to which users interact with a digital library depends on how well users understand the terminology displayed on the system interface. Talja et al. (1998) have found that the issue of inappropriately used jargons is one of the major problems with digital libraries. The problems reported in library literature include ambiguity of terminology and the fact that librarians/designers assume too much about users' knowledge. McGillis and Toms (2001), for example, state that librarians make too many assumptions about the extent of user knowledge. McDonald (2001) reported that users clearly do not understand the library meaning of reference and index although those are familiar to librarians. He stated that, from a user standpoint, the term index could be used to mean the index of a
book volume but was not clearly understood when used to indicate a source that indexes journals and periodical content. Duncker (2002) reported that students understood the terms *title* and *full-text*, but the terms *reviews* and *index terms* were not known. Gullikson et al. (1999) indicated that users were clearly confused with the inclusiveness of some terms. Detlor et al. (2003) reported that there was some confusion by lay users with terms like *hold* or *reserves* in that some users think of "reserving" library books rather than placing a hold on them. King (2001) reported that the Kansas City Public Library has changed its Web site (www.kclibrary.org) wording to more customer-focused terms rather than using wording created by and for librarians. Fox et al. (1993) reported that the wording of messages on the display was changed following suggestions from the usability test results. Hartson, Shivakumar, Pérez-Quiñones (2004) reported from their usability inspection of NCSTRL that problems with wording accounted for 36% of the problems. Although this issue may be seen as a library instruction issue that can be overcome by better user education, better labeling in the digital library is needed to enhance usability. As Hartson, Shivakumar, Pérez-Quiñones (2004) put it, "precise use of words in user interfaces is one of the utmost important design considerations for usability" (p. 112).

Walbridge (2000) indicated that users have trouble with concepts, such as, what is in the catalog, and what differentiates the catalog from those in electronic databases. This is not a terminology issue or the fault of the system's being designed from a librarian's perspective, but rather an issue of user education.

The usability testing at Roger Williams University and Western Michigan University indicated that finding periodical articles and subject specific databases was a
challenge for users (Cockrell & Jayne, 2002; McMullen, 2001a; McMullen, 2001b). The pilot study of the present research confirmed this observation. Therefore, the present study includes in the research instruments the tools to evaluate the system's ease of use in locating journals, articles, encyclopedia, and e-books.

Kengeri et al. (1999) studied usability from the perspectives of effectiveness, likeability, learnability, and usefulness. Subjects were grouped into experienced and novice users to study four digital libraries: ACM (http://www.acm.org), IEEE-CS (http://www.computer.org), NCSTRL (http://www.ncstrl.org), and NDLTD (http://www.theses.org). The study found that the novices and the experienced users had similar performances. However, the study of Alexander (2003) found that expert users generally completed the tasks in less time than novices, while the number of tasks completed correctly and the number of usability problems reported did not differ significantly by experience level.

The study at Washington State University (Chisman, Diller, & Walbridge, 1999) found that users are so supremely confident in their searching that when they do not find something, they immediately conclude that the Libraries do not own the item. It does not occur to them that they may have searched incorrectly and that they could or should try another search.

Sumner and Dawe (2001) examined the usability of the Digital Library for Earth System Education (DLESE) and reported that there was a problem of mismatch between metadata and the resource displayed. They further reported that the design of the search result page is critical. There must be enough information to effectively summarize the resource, but not so much text that people skip reading it.
Collins and Aguiñaga (2001), using 35 subjects, suspect that there are gender, age, and cultural differences in the ways people interact with online information, although they draw no firm conclusions. Vohringer-Kuhnt (2003), using 145 subjects from 30 different countries, found that there are differences in attitude towards usability among members of different national groups. Duncker, Theng, and Mohd-Nasir (2000) studied 60 students at Middlesex University, and found that students with different cultural backgrounds have different color preferences. Duncker (2002) studied Māori library users and found that Western libraries are counterintuitive for them. Borgman and Rasmussen (2005) observe that “DLs have made tremendous progress in making information available across borders and cultures, but the usability of these systems in multicultural and cross-cultural environments is only beginning to be examined” (p. 274). In addition, the usability study at the University of the Pacific showed significant differences among certain groups of academic majors or fields of study (Krueger, Ray, & Knight, 2004). Those demographic factors will be examined in the present study.

Academic library Web sites serve a wide range of users in terms of age, gender, discipline, and culture. It should be interesting to observe if these variables influence usability assessment.

An interesting observation is made by McGillis and Toms (2001) about the click cost phenomenon: that users are reluctant to click a link unless they are fairly certain that they will discover what they are looking for. This has been briefly discussed in the Chi-Web listserv but not formally investigated. The present dissertation study will examine this issue in the experiment.

In summary, the key points of concern are:
• user lostness
• navigation disorientation
• lack of benchmarks
• design from librarian’s/designer’s perspective instead of from user’s
• cross-culture usability
• ambiguity of terminology
• click cost

These issues synthesized from the literature review will be carefully examined in the present study to see if the two test sites have similar problems.
Chapter 3

Usability Evaluation Model

Based on the literature review presented in chapter 2, a model of usability evaluation for the academic digital library is presented in this chapter.

Usability is a multidimensional construct and can be assessed using various criteria, as discussed in Section 2.2. This research applies the definition of ISO 9241-11 (1994) that examines effectiveness, efficiency, and satisfaction. In addition, this research also examines learnability. The ISO definition covers the performance criteria (effectiveness and efficiency) as well as how users like the system. The learnability criterion is recommended by usability expert Nielsen (1993) as the most fundamental usability criterion. The examination of effectiveness, efficiency, satisfaction (which is further examined by ease of use, organization of information, labeling and terminology, visual appearance, content, and error correction), and learnability make the evaluation model a well-rounded one. A digital library or an information system should be designed to allow users to get to the desired information quickly, easily, and accurately.

The ISO (1994) defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (p. 10). The ISO definition, however, does not specify operational criteria for evaluation. Schwartz (2000) found that "[a]lthough measuring efficiency might be relatively straightforward, there is little agreement as to how to measure effectiveness and satisfaction in library contexts" (p. 390). The present dissertation provides operational criteria for this purpose.
In addition, Karoulis and Pombortsis (2003) suspect that usability (effectiveness, efficiency, and satisfaction) and learnability of educational environment are positively correlated and wonder how far one affects the other. Karoulis and Pombortsis never actually carried out a study to examine this possible correlation, nor did they provide operational criteria. The present dissertation will study this relationship.

In this study, to evaluate effectiveness is to evaluate whether the system as a whole can provide information and functionality effectively; it will be measured by the number of correct answers. To evaluate efficiency is to evaluate whether the system as a whole can be used to retrieve information efficiently; it will be measured by the length of time it takes to complete tasks and the number of steps required. In evaluating satisfaction we will look into the areas of ease of use, organization of information, clear labeling, visual appearance, contents, and error correction; it will be measured by Likert scales and questionnaires. Belkin and Vickery (1985) have stated, “Although the idea behind satisfaction as a criterion is simple, it is obvious that actually measuring it is not, since it must be a multifaceted variable” (p. 195). This study assesses satisfaction from the perspectives of the users’ reactions to ease of use, organization of information, clear labeling, visual appearance, contents, and error corrections. To determine ease of use we will evaluate the users’ perceptions of the ease of use of the system. To evaluate the organization of information we will evaluate whether the system’s structure, layout, and organization meets the users’ satisfaction. To evaluate the labeling we will explore the users’ opinions as to whether the system provides clear labeling and whether the terminology used is easy to understand. The users’ evaluations of the site’s design will provide the ratings for visual appearance. To evaluate the contents we will evaluate the
authority and accuracy of the information provided. To evaluate error we will test whether or not users are able to recover from mistakes easily and whether they make mistakes easily due to the system's design. To evaluate learnability we will measure the learning effort required; it can be examined by asking subjects to cross search the other test site: that is, Rutgers participants will search the Queens site and vice versa. It is then measured by (a) how soon the subjects know how to begin searching, (b) how many tasks are completed correctly, and (c) how much time it takes to complete the tasks.

Figure 2 is a diagram illustrating this evaluation model. It is suspected that there exists an interlocking relationship among effectiveness, efficiency, and satisfaction. It is also interesting to examine how learnability interacts with these three factors.

This evaluation model considers both quantifying elements of performance (time, accuracy rate, steps to complete tasks) and subjective criteria (satisfaction). In addition, the instruments seek opinions from participants regarding overall reaction, best features, worst features, desired features, and suggestion/comments. The model and instruments evaluate usability wholly, both inherent and apparent, covering outcome, process, and task attributes.

The evaluation approach is empirical, exploratory, and user-centered.

Table 3 outlines each factor proposed in the evaluation model, including definition, criteria, and measure.
This study examines four important criteria of usability – effectiveness, efficiency, satisfaction, and learnability. It will be interesting to examine whether or not there exists an interlocking relationship.
Table 3  
Factors of the Evaluation Model

**Effectiveness:**

| Definition: | the ability of the system as a whole to provide information and function effectively |
| Criteria:   | correctness/incorrectness |
| Measure:    | yes/no |

**Efficiency:**

| Definition: | the ability of the system as a whole to retrieve correct information in a timely and straightforward manner |
| Criterion:  | length of time to complete tasks |
| Measure:    | minutes/seconds |

**Satisfaction:**

| Definition: | the users' reaction to the system |
| Criteria:   | ease of use, organization of information, clear labeling, visual appearance, contents, and error correction |
| Measure:    | 5-point Likert scale |

**Ease of Use:**

| Definition: | the susceptibility of the system to simple manipulation |

**Organization of Information:**

| Definition: | the system's structure, layout, and information organization |

**Clear Labeling:**

| Definition: | visual and conceptual clarity of labeling |

**Visual Appearance:**

| Definition: | visual attractiveness of site |

**Content:**

| Definition: | authority and accuracy of information provided |

**Error Correction:**

| Definition: | ability to recover quickly and easily from mistakes |
Learnability:

**Definition:** learning effort required to use a new site

**Criteria:**
1) length of time to begin searching
2) number of questions answered correctly
3) length of time taken to answer questions

**Measure:**
1) seconds
2) yes/no
3) minutes/seconds
Chapter 4

Methods of Data Collection

In this chapter the method of the study is described including the selections of sites and participants, stages, tasks, materials, setting, procedure, and pilot study.

4.1 Sites

The Rutgers University Libraries Web site (http://www.libraries.rutgers.edu) and the Queens College Web site (http://qcpages.qc.edu/Library/) were selected to test the instruments designed in this dissertation. These two sites provide rich resources in their local holdings, connections to electronic resources and digital collections, and are the same type of digital library. The use of two test sites provides an opportunity to evaluate learnability and cross-institutional usability.

4.2 Stages

The study was divided into two stages: stage one was conducted in February/March of 2004 and stage two was conducted in September/October of 2004. The Rutgers University Libraries revamped its Web site in September 2004, which in part prompted an interest in repeating the usability testing. The other reason for re-running the usability testing was to confirm the interlocking relationship among effectiveness, efficiency, and satisfaction that was discovered in stage one.
4.3 Subjects

This study was conducted in two stages. The first stage had a total of eleven subjects (five from Rutgers and six from Queens). The second stage had a total of thirty subjects (fifteen from Rutgers and fifteen from Queens). Therefore, the sum of subjects for this study was forty-one subjects, composed of both graduate and undergraduate students. The convenience sampling method was used for recruitment. The literature review indicates that the convenience sampling method is common in usability testing. Rutgers students were recruited by announcing the usability tests on various students’ listservs. Queens students were recruited by announcing the usability tests in classes.

4.4 Tasks

This study applied the formal usability testing technique to evaluate the usability of the Rutgers and Queens libraries’ Web sites. Formal usability testing has five characteristics: (a) The primary goal is to improve the usability of a product; (b) The participants represent real users; (c) The participants do real tasks; (d) The participants’ actions and words are observed and recorded; and (e) The data are analyzed, the real problems are diagnosed, and recommendations to fix those problems are made (Dumas & Redish, 1993). It is an empirical, exploratory, user-centered approach.

A list of nine tasks was designed to test the effectiveness, efficiency, and user satisfaction of the system (see Appendix C). The nine tasks were designed to be representative of typical uses of a library’s Web site, and were as follows:

1. Does the library have a copy of Gone with the Wind, book format, by Margaret Mitchell?
2. Does the library currently subscribe to paper copy of Advertising Age?
3. Use a database to find an article about nursing homes and mental illness.
4. Use a database to find a journal article on gospel music.
5. I am interested in investing in what are referred to as “callable securities.”
   Please use a database to find a recent article about them.
6. Find an encyclopedia article about French wine.
7. Find an e-book called The story of mankind.
8. Can alumni enjoy inter-library loan service?
9. Find instructions on setting up your home computer to have remote access to the library’s electronic resources.

Tasks 1, 2, and 7 were to find known items; tasks 3, 4, 5, and 6 were to find articles about various topics; tasks 8 and 9 were to locate information. These are representative tasks when using an academic library Web site. Each subject was asked to search two sites with the sequence alternated. When searching the other institution’s site, tasks 3-5 were eliminated because of proxy server authorization and to limit the session to a reasonable timeframe. The purpose of searching the other institution’s site was to measure learnability.

4.5 Instruments

A packet of materials was presented to each subject, including a consent form (see Appendixes A-1 and A-2), a pre-search questionnaire (see Appendix B), two task lists (see Appendix C), and two copies of post-search questionnaire (see Appendixes D-1 and
The consent form was to request the subject's permission to record his/her actions. The pre-search questionnaire was used to collect demographic data and an idea of the subject's familiarity with the Web sites. The post-search questionnaire was designed to test the subject's satisfaction with the site and to gather comments. It was revised after stage one, based on the observations. Two copies of Appendixes C and D were needed for the subjects to conduct searches on two test sites. When the subjects were asked to search another institution's site, the tasks of searching databases (i.e., tasks 3, 4, 5) were eliminated, partly due to proxy server authorization and partly to limit the session to less than 1 ½ hours.

4.6 Setting

The usability tests were conducted (a) in the Usability Laboratory (Room 214) at the School of Communication, Information, and Library Studies at Rutgers University, for Rutgers subjects; and (b) in a faculty office, for Queens subjects.

The Rutgers SCILS Usability Laboratory has two rooms, one for conducting testing and one for observation, divided by a one-way mirror. A ceiling-mounted video camera recorded the entire session with a focus on the screen activities. The subject's voice and a time stamp were on the tape. In addition, transaction log software named TrueActive was used to record transactions. The experimenter was in the testing room to answer questions and to conduct post-search interviews.

The study for Queens subjects was conducted in a faculty office. The transactions were recorded by camcorder and TrueActive transaction log software.
4.7 Procedure

Upon arrival, the subject signed the consent form, answered the pre-search questionnaire, and received a list of nine tasks. The computer was set to have the RUL Web site or the Queens College Library Web site on the screen, depending on the predetermined sequence for the subject. Half of the subjects searched the Rutgers site first and the other half searched the Queens site first. Each subject performed the nine tasks at his/her pace and was encouraged to think aloud. After each task was performed, the computer screen was reset to the home page of the site to allow for a fresh start for the next task. The screen actions and the subject’s verbal protocol were videotaped. The transactions were further recorded by transaction log software. At the end, the subjects answered a post-search questionnaire. The conversation was recorded on audiotape. The subjects then performed the same tasks, excluding the three database search questions, on the other test site. The session took about one hour and a half.

4.8 Pilot Study

A pilot study of three subjects at Rutgers was conducted in April 2002. The level of difficulty of the tasks as well as the post-test questionnaire were analyzed and revised. The results of the pilot study are not reported in this dissertation.
Chapter 5

Data Analysis

This chapter reports the results of both stages of the studies: stage one of eleven subjects (five from Rutgers and six from Queens), which took place in February/March 2004, and stage two of thirty subjects (fifteen from Rutgers and fifteen from Queens), which took place in September/October 2004. Subjects were numbered by institution, stage, and then number. For example, R1-2 denotes Rutgers subject 2 in stage one; Q2-1 denotes Queens subject 1 in stage two. For the tables illustrated in this dissertation, $N_{R1}$ denotes the number of Rutgers subjects in stage one; $N_{Q2}$ denotes the number of Queens subjects in stage two. A summary of findings is available at the end of this chapter.

5.1 Demographic Distribution

5.1.a. Stage One:

There were five subjects from Rutgers in stage one. One was male, four were female. One was an undergraduate, two were master’s degree students, and two were doctoral students. Age range was from 22 to 31. One was in the field of communication, three were in library and information science, and one was in neuroscience. They had been at Rutgers from 1 month to 4 years. Two were White, two were Asian, and one was African-American. In terms of how often they had used the Rutgers Web site, two said they had used it once or twice a month, two said once or twice a week, and one said daily. None of them had used the Queens Web site.

The six subjects at Queens College included three males and three females; all in the master’s program of library and information science; four were White, one was
Asian-American, and one was African-American; years at Queens ranged from one to six years. One was in his/her 20s, three in their 30s, and two in their 50s. As to their experience with the Queens College Web site, one used it once or twice a month, three used it once or twice a week, and two used it on a daily basis. None had used the Rutgers Libraries Web site.

5.1.b. Stage Two:

In stage two of the study, there were a total of thirty subjects: fifteen from Rutgers and fifteen from Queens.

The Rutgers subjects comprised four males and eleven females; eight undergraduates, three master’s degree students, and four doctoral students. Two were in their teens, eight in their 20s, two in their 30s, and three in their 40s. Seven majored in library and information science and eight in communication. They had been at Rutgers less than 1 year (1), 1 year (2), 2 years (5), 3 years (5), and 4 years (2). Of the four who came from foreign countries, one had come to the U.S. at age 2 and the other three had been in the U.S. for less than 5 years. One was from Dominica, two from Korea, and one from Taiwan. As to ethnic distribution, seven were White, two were African-Americans, five were Asian-Americans, and one was Hispanic. Their experience in using library Web sites ranged from daily (2), to once or twice a week (7), to once or twice a month (2), to once or twice a semester (4).

The Queens subjects comprised seven males and eight females, all in the master’s program of library and information science. Two were in their 20s, six in their 30s, three in their 40s, and four in their 50s. They had been at Queens College for 1 year (7), 2
years (7), and 8 years (1). Two were from Hong Kong; the others were born and raised in the U.S. Twelve were White, one was African-American, and two were Asian. Their experience in using library Web sites ranged from daily (9), to once or twice a week (4), to once or twice a month (1), to once or twice a semester (1).

Table 4 tabulates the participants’ demographic data.

Table 4

Participants Demographic Data (N = 41)

<table>
<thead>
<tr>
<th></th>
<th>Stage 1</th>
<th></th>
<th>Stage 2</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rutgers</td>
<td>Queens</td>
<td>Rutgers</td>
<td>Queens</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>Female</td>
<td>4</td>
<td>3</td>
<td>11</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>20-29</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>30-39</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>40-49</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>&gt;=50</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Master’s student</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>Doctoral student</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Major</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIS</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>Communication</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Neuroscience</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Years at the Institution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>less than one year</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>one year</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>two years</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>three years</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>four years</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>more than four years</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td>African or African-American</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Asian or Asian American</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Frequency of Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Once or twice a semester</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Once or twice a month</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Once or twice a week</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>Daily</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>14</td>
</tr>
</tbody>
</table>
5.2 Effectiveness

Effectiveness in this study is measured by how many answers are correct. Each subject was asked to perform fifteen tasks, including nine at his or her library site (see Appendix C) and six at the other library site. The sequence was alternated: one half started with the Rutgers site and the other half started with the Queens site. Because of proxy server authorization and to limit the session to less than an hour and a half, tasks 3 to 5 in Appendix C were eliminated when searching the other library's site.

Table 5 displays the effectiveness scores, as expressed in percentages of correctness, of Rutgers subjects searching the Rutgers site and Queens subjects searching the Queens site respectively. Table 6 shows the performances of all eleven subjects in stage one and all thirty subjects in stage two, in searching the Rutgers site and the Queens site respectively.

Table 5

<table>
<thead>
<tr>
<th>Effectiveness of Rutgers Subjects Searching Rutgers Site and Queens Subjects Searching Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Task 1</td>
</tr>
<tr>
<td>Task 2</td>
</tr>
<tr>
<td>Task 3</td>
</tr>
<tr>
<td>Task 4</td>
</tr>
<tr>
<td>Task 5</td>
</tr>
<tr>
<td>Task 6</td>
</tr>
<tr>
<td>Task 7</td>
</tr>
<tr>
<td>Task 8</td>
</tr>
<tr>
<td>Task 9</td>
</tr>
<tr>
<td>Average</td>
</tr>
</tbody>
</table>
Table 6
Effectiveness of All Subjects Searching Rutgers and Queens Sites

<table>
<thead>
<tr>
<th>Task</th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td></td>
<td>N = 11</td>
<td>N = 30</td>
</tr>
<tr>
<td>Task 1</td>
<td>Find a book</td>
<td>73%</td>
</tr>
<tr>
<td>Task 2</td>
<td>Find a journal</td>
<td>91%</td>
</tr>
<tr>
<td>Task 3</td>
<td>Use database</td>
<td>80%</td>
</tr>
<tr>
<td>Task 4</td>
<td>Use database</td>
<td>100%</td>
</tr>
<tr>
<td>Task 5</td>
<td>Use database</td>
<td>100%</td>
</tr>
<tr>
<td>Task 6</td>
<td>Find encyclopedia</td>
<td>55%</td>
</tr>
<tr>
<td>Task 7</td>
<td>Find e-book</td>
<td>64%</td>
</tr>
<tr>
<td>Task 8</td>
<td>Find information</td>
<td>100%</td>
</tr>
<tr>
<td>Task 9</td>
<td>Find information</td>
<td>91%</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>84%</td>
</tr>
</tbody>
</table>

As indicated in Tables 5 and 6, task 6 (find an encyclopedia article) and task 7 (find an e-book) gave subjects more difficulties on both sites. This is consistent in stages one and two. It is also consistent when comparing the scores of subjects searching their own institutions and the scores of all subjects. It was difficult in stage one to dig out e-book collections in Rutgers’ cluster of information. However, it is surprising that subjects still have difficulty in locating e-books at the Rutgers site when their MARC records became readily searchable on Rutgers’ online catalog in stage two. The Queens site promotes e-books on its home page (see Figure 3). However, the NetLibrary collection at the Queens site was not listed under e-books but was grouped with online catalogs (see Figures 4 and 5). Furthermore, many subjects did not recognize the NetLibrary icon as the logon to the NetLibrary collection but thought the NetLibrary page was just a description page (see Figure 6).
Figure 3. Queens College Library Web Page in June 2004
Figure 4. Queens College Library Web Site in June 2004, Library Catalogs
Figure 5. Queens College Library Web Site in June 2004, E-Books
5.3 Efficiency

Efficiency in this study is measured by the length of time it took to complete a task correctly and the number of keystrokes/clicks (or steps/movements) required. If the answer is wrong, the time spent and the steps were not counted in the calculation. If a subject decided to give up a task, the time spent and the steps taken on the particular task were not counted. Table 7 indicates time used by Rutgers subjects to search the Rutgers site and by Queens subjects to search the Queens site on stages one and two respectively. Table 8 tabulates the time used by all eleven subjects in stage one and all thirty subjects in stage two respectively.
Table 7

*Time Used by Subjects to Search Their Own Institutions’ Sites*

<table>
<thead>
<tr>
<th></th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td></td>
<td>$N_{R1}=5$</td>
<td>$N_{R2}=15$</td>
</tr>
<tr>
<td>Task 1</td>
<td>Find a book</td>
<td>3 min. 17 sec.</td>
</tr>
<tr>
<td>Task 2</td>
<td>Find a journal</td>
<td>1 min. 9 sec.</td>
</tr>
<tr>
<td>Task 3</td>
<td>Use database</td>
<td>6 min. 11 sec.</td>
</tr>
<tr>
<td>Task 4</td>
<td>Use database</td>
<td>3 min. 18 sec.</td>
</tr>
<tr>
<td>Task 5</td>
<td>Use database</td>
<td>6 min. 33 sec.</td>
</tr>
<tr>
<td>Task 6</td>
<td>Find encyclopedia</td>
<td>4 min. 12 sec.</td>
</tr>
<tr>
<td>Task 7</td>
<td>Find e-book</td>
<td>6 min. 17 sec.</td>
</tr>
<tr>
<td>Task 8</td>
<td>Find information</td>
<td>2 min. 4 sec.</td>
</tr>
<tr>
<td>Task 9</td>
<td>Find information</td>
<td>1 min. 23 sec.</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td>3 min. 49 sec.</td>
</tr>
</tbody>
</table>

Table 8

*Time Used by All Subjects in Stages One and Two on Both Sites*

<table>
<thead>
<tr>
<th></th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td></td>
<td>$N = 11$</td>
<td>$N = 30$</td>
</tr>
<tr>
<td>Task 1</td>
<td>Find a book</td>
<td>2 min. 40 sec.</td>
</tr>
<tr>
<td>Task 2</td>
<td>Find a journal</td>
<td>59 sec.</td>
</tr>
<tr>
<td>Task 3</td>
<td>Use database</td>
<td>6 min. 11 sec.</td>
</tr>
<tr>
<td>Task 4</td>
<td>Use database</td>
<td>3 min. 18 sec.</td>
</tr>
<tr>
<td>Task 5</td>
<td>Use database</td>
<td>6 min. 33 sec.</td>
</tr>
<tr>
<td>Task 6</td>
<td>Find encyclopedia</td>
<td>3 min. 32 sec.</td>
</tr>
<tr>
<td>Task 7</td>
<td>Find e-book</td>
<td>5 min. 28 sec.</td>
</tr>
<tr>
<td>Task 8</td>
<td>Find information</td>
<td>1 min. 16 sec.</td>
</tr>
<tr>
<td>Task 9</td>
<td>Find information</td>
<td>59 sec.</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td>3 min. 26 sec.</td>
</tr>
</tbody>
</table>

As indicated in Tables 7 and 8, task 9 (find information to set up remote access) is the easiest on both Rutgers and Queens sites. This reflects that these two sites recognize the importance of library users’ being able to remotely access a library’s electronic resources easily and thus make this task easy. Although the instructions are readily

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
available and easy to locate, it is equally important that they be easy to follow. This study does not evaluate the clarity or ease of use of the instructions.

In addition to using time as an indicator to assess efficiency, the number of steps required is also considered in this study. This is counted by the number of keystrokes or clicks. A string of characters when performing a title or author search is counted as one step. Each press on the Enter key or each click on the mouse is also counted as one step. Table 9 indicates the numbers of steps taken by Rutgers and Queens subjects in searching their own sites. Table 10 indicates the number of steps for all subjects on these two sites.

Table 9

Number of Steps Taken by Rutgers Subjects to Search Rutgers Site and Queens Subjects to Search Queens Site

<table>
<thead>
<tr>
<th>Task</th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Task 1</td>
<td>Find a book</td>
<td>9</td>
</tr>
<tr>
<td>Task 2</td>
<td>Find a journal</td>
<td>5</td>
</tr>
<tr>
<td>Task 3</td>
<td>Use database</td>
<td>20</td>
</tr>
<tr>
<td>Task 4</td>
<td>Use database</td>
<td>9</td>
</tr>
<tr>
<td>Task 5</td>
<td>Use database</td>
<td>19</td>
</tr>
<tr>
<td>Task 6</td>
<td>Find encyclopedia</td>
<td>15</td>
</tr>
<tr>
<td>Task 7</td>
<td>Find e-book</td>
<td>27</td>
</tr>
<tr>
<td>Task 8</td>
<td>Find information</td>
<td>6</td>
</tr>
<tr>
<td>Task 9</td>
<td>Find information</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>13</td>
<td>9</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 10

Number of Steps Taken by All Subjects in Stages One and Two on Both Sites

<table>
<thead>
<tr>
<th>Task</th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Task 1</td>
<td>Find a book</td>
<td>7</td>
</tr>
<tr>
<td>Task 2</td>
<td>Find a journal</td>
<td>5</td>
</tr>
<tr>
<td>Task 3</td>
<td>Use database</td>
<td>20</td>
</tr>
<tr>
<td>Task 4</td>
<td>Use database</td>
<td>9</td>
</tr>
<tr>
<td>Task 5</td>
<td>Use database</td>
<td>19</td>
</tr>
<tr>
<td>Task 6</td>
<td>Find encyclopedia</td>
<td>11</td>
</tr>
<tr>
<td>Task 7</td>
<td>Find e-book</td>
<td>24</td>
</tr>
<tr>
<td>Task 8</td>
<td>Find information</td>
<td>4</td>
</tr>
<tr>
<td>Task 9</td>
<td>Find information</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>11</td>
</tr>
</tbody>
</table>

As indicated in Tables 9 and 10, task 7 (find e-book) took subjects more steps to answer on the Rutgers site in both stages one and two, by Rutgers subjects only and by all subjects. Three participants (R2-6, R2-7, and R2-9) suggested that Rutgers should make e-book collections more noticeable on its Web page. It was hard to dig out e-book collections in the cluster of the Rutgers site. It is good news to find that the MARC records of NetLibrary became available as part of the Rutgers online catalog on October 18, 2004. However, this was near the end of stage two experiments when all the Rutgers subjects had completed their sessions and there were only four Queens subjects to go.

Task 6 (find encyclopedia article) required more steps on the Queens site in both stages one and two, by Queens subjects only and by all subjects. R2-12 said it was very difficult to find an encyclopedia on the Queens site: “Better descriptions necessary.”

The use of electronic databases to find articles was not easy for participants. Four participants (R2-6, R2-11, R2-12, and R2-13) commented that it was necessary to know those databases in order to select one from the long list. Two participants (R2-2 and R2-
9) said that they had difficulty in combining terms for task 3 (find an article about nursing homes and mental illness). Two participants (R2-2 and R2-15) indicated that the database subject guides on the Rutgers site were helpful. Two participants (R1-5 and R2-13) suggested meta-searching as a solution. In that case, the user enters search strings to search all available databases and does not need to learn various interfaces and commands. R2-15 also felt that searching databases was confusing although the subject guides on the Rutgers site were helpful.

For task 1 (find a book), the Queens site was easier to use. This is reflected in a higher accuracy rate, shorter time, and fewer clicks. It is also reflected in the satisfaction score reported in the later section of this dissertation. R2-3, for instance, commented that the Queens online catalog was surprisingly easier to use. Several subjects commented that the search results display of the Rutgers site was too complicated and the searches resulted in too many hits. More comments related to task 1 are discussed in the satisfaction section later in this dissertation.

Task 2 was about a paper subscription to a journal. Several subjects had difficulty in distinguishing among the terms paper copy, microform copy, and electronic copy.

On both sites it was very easy for subjects to find instructions on setting up remote access from home. The Rutgers site places “Connect from home” on the left navigation pane on its home page, while the Queens site places it as the first entry of “How do I ...?”
5.4 Learnability

Learnability is in some sense the most fundamental usability attribute (Nielsen, 1993). The system should be easy to learn so that the user can rapidly start getting some work done with it. As Dzida, Herda, and Itzfeldt (1978) report, learnability is especially important for novice users.

This study examined the learning effort involved in using a new Web site and measured how easy it was (or was not) for new visitors to orient themselves and get a good overview of what the site offered. Learnability was inferred from the amount of time required for new users to achieve user performance standards and perform tasks correctly. The design of this study required Rutgers subjects to search the Queens site and Queens subjects to search the Rutgers site. It found that all subjects could begin searching without difficulty and began to perform searches almost immediately. The experimenter counted the time subjects took before their first click and found that they started almost immediately.

Learnability was also examined by the number of questions subjects could answer correctly and the length of time it took on the new Web site. This was to determine if they could perform tasks at a proficient level. Table 11 shows the results by correctness of answers in stage one and Table 12 in stage two. As stated earlier, subjects performed a total of nine tasks on their home site but only six on the new site. Tasks 3 to 5 were eliminated when working on the new site because of proxy server authorization and to limit the session to a specific time frame.

The results in stage one indicate that the Rutgers subjects achieved 83% correctness when using the Rutgers Web site, while reaching 90% correctness when...
using the Queens Web site. However, the Queens subjects achieved the same level of correctness, 75%, on both the Rutgers and the Queens College Web sites. This might suggest that the Queens site offered higher learnability.

However, the results in stage two indicate that the Rutgers subjects achieved 77% correctness when using the Rutgers site while reaching only 71% correctness when using the Queens site, while the Queens subjects achieved 67% on the Rutgers site but 81% on their home site. This may be explained by the possibility that subjects felt more comfortable with their home site than with a new site.

Table 11

**Learnability Comparison by Correctness of Answers in Stage One**

<table>
<thead>
<tr>
<th>Task</th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rutgers subjects</td>
<td>Queens subjects</td>
</tr>
<tr>
<td></td>
<td>$N_{Rt}=5$</td>
<td>$N_{Qt}=6$</td>
</tr>
<tr>
<td>Task 1</td>
<td>80%</td>
<td>67%</td>
</tr>
<tr>
<td>Task 2</td>
<td>100%</td>
<td>83%</td>
</tr>
<tr>
<td>Task 6</td>
<td>40%</td>
<td>67%</td>
</tr>
<tr>
<td>Task 7</td>
<td>80%</td>
<td>50%</td>
</tr>
<tr>
<td>Task 8</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Task 9</td>
<td>100%</td>
<td>83%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>83%</strong></td>
<td><strong>75%</strong></td>
</tr>
</tbody>
</table>

Table 12

**Learnability Comparison by Correctness of Answers in Stage Two**

<table>
<thead>
<tr>
<th>Task</th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rutgers subjects</td>
<td>Queens subjects</td>
</tr>
<tr>
<td></td>
<td>$N_{Rt}=15$</td>
<td>$N_{Qt}=15$</td>
</tr>
<tr>
<td>Task 1</td>
<td>100%</td>
<td>73%</td>
</tr>
<tr>
<td>Task 2</td>
<td>80%</td>
<td>73%</td>
</tr>
<tr>
<td>Task 6</td>
<td>53%</td>
<td>20%</td>
</tr>
<tr>
<td>Task 7</td>
<td>33%</td>
<td>33%</td>
</tr>
<tr>
<td>Task 8</td>
<td>93%</td>
<td>100%</td>
</tr>
<tr>
<td>Task 9</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>77%</strong></td>
<td><strong>67%</strong></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Learnability was also measured by how much time the subjects took to complete tasks on the new site as compared with the time they took on their home institution’s site. Table 13 indicates that both Rutgers and Queens subjects were able to complete tasks in less time on the Queens site but took more time on the Rutgers site in stage one. When looking at the information on Tables 11 and 13, we find that subjects were able to achieve a higher accuracy rate in less time on the Queens site than on the Rutgers site.

Table 14 is a similar comparison for stage two. As indicated, Rutgers subjects completed tasks using the same amount of time on both sites but Queens subjects were able to use less time on their home site but took more time on the Rutgers site.

Table 13

Learnability by Time in Stage One

<table>
<thead>
<tr>
<th>Task</th>
<th>Rutgers Web site</th>
<th>Queens Web site</th>
<th>Rutgers Web site</th>
<th>Queens Web site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rutgers subjects</td>
<td>Queens subjects</td>
<td>Rutgers subjects</td>
<td>Queens subjects</td>
</tr>
<tr>
<td></td>
<td>(N_{R_1}=5)</td>
<td>(N_{Q_1}=6)</td>
<td>(N_{R_1}=5)</td>
<td>(N_{Q_1}=6)</td>
</tr>
<tr>
<td>Task 1</td>
<td>3 min. 17 sec.</td>
<td>2 min. 3 sec.</td>
<td>1 min. 58 sec.</td>
<td>1 min. 29 sec.</td>
</tr>
<tr>
<td>Task 2</td>
<td>1 min. 9 sec.</td>
<td>50 sec.</td>
<td>2 min. 47 sec.</td>
<td>1 min. 5 sec.</td>
</tr>
<tr>
<td>Task 6</td>
<td>4 min. 12 sec.</td>
<td>3 min. 11 sec.</td>
<td>2 min. 52 sec.</td>
<td>2 min. 20 sec.</td>
</tr>
<tr>
<td>Task 7</td>
<td>6 min. 17 sec.</td>
<td>4 min. 39 sec.</td>
<td>3 min. 31 sec.</td>
<td>1 min. 36 sec.</td>
</tr>
<tr>
<td>Task 8</td>
<td>2 min. 4 sec.</td>
<td>35 sec.</td>
<td>1 min. 46 sec.</td>
<td>1 min. 8 sec.</td>
</tr>
<tr>
<td>Task 9</td>
<td>1 min. 23 sec.</td>
<td>47 sec.</td>
<td>1 min. 7 sec.</td>
<td>1 min. 10 sec.</td>
</tr>
<tr>
<td>Average</td>
<td>3 min. 4 sec.</td>
<td>2 min.</td>
<td>2 min. 20 sec.</td>
<td>1 min. 42 sec.</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 14

Learnability by Time in Stage Two

<table>
<thead>
<tr>
<th></th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subjects</td>
<td>Subjects</td>
</tr>
<tr>
<td>$N_{R}=15$</td>
<td>$N_{Q}=15$</td>
<td>$N_{R}=15$</td>
</tr>
<tr>
<td>Task 1</td>
<td>2 min. 12 sec.</td>
<td>3 min. 32 sec.</td>
</tr>
<tr>
<td></td>
<td>1 min. 53 sec.</td>
<td>1 min. 28 sec.</td>
</tr>
<tr>
<td>Task 2</td>
<td>1 min. 39 sec.</td>
<td>1 min. 17 sec.</td>
</tr>
<tr>
<td></td>
<td>2 min. 21 sec.</td>
<td>1 min. 26 sec.</td>
</tr>
<tr>
<td>Task 6</td>
<td>2 min. 50 sec.</td>
<td>3 min. 42 sec.</td>
</tr>
<tr>
<td></td>
<td>4 min. 15 sec.</td>
<td>4 min. 33 sec.</td>
</tr>
<tr>
<td>Task 7</td>
<td>5 min. 25 sec.</td>
<td>3 min. 41 sec.</td>
</tr>
<tr>
<td></td>
<td>3 min. 20 sec.</td>
<td>2 min. 11 sec.</td>
</tr>
<tr>
<td>Task 8</td>
<td>46 sec.</td>
<td>57 sec.</td>
</tr>
<tr>
<td></td>
<td>1 min. 3 sec.</td>
<td>1 min. 9 sec.</td>
</tr>
<tr>
<td>Task 9</td>
<td>29 sec.</td>
<td>37 sec.</td>
</tr>
<tr>
<td></td>
<td>33 sec.</td>
<td>46 sec.</td>
</tr>
<tr>
<td>Average</td>
<td>2 min. 14 sec.</td>
<td>2 min. 18 sec.</td>
</tr>
</tbody>
</table>

Nielsen (1993) suggests that learnability can be measured by picking some users who never used the system before and measuring the time it takes them to reach a specified level of proficiency in using it. The most common way to express the specified level of proficiency is simply to state that the users have to be able to complete a certain task successfully. The present study considers three measures to examine learnability: (a) how much time it takes to begin to perform a task, (b) the rate of correctness, and (c) the time spent on completing tasks correctly. Tables 15 and 16 are summary tables for effectiveness and efficiency in the two stages respectively.

Table 15

Learnability of Both Sites in Stage One

<table>
<thead>
<tr>
<th></th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subjects</td>
<td>Subjects</td>
</tr>
<tr>
<td>$N_{R}=5$</td>
<td></td>
<td>$N_{Q}=6$</td>
</tr>
<tr>
<td>Correctness</td>
<td>83%</td>
<td>90%</td>
</tr>
<tr>
<td>Speed</td>
<td>3 min. 4 sec.</td>
<td>2 min. 20 sec.</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>75%</td>
</tr>
<tr>
<td></td>
<td>2 min.</td>
<td>1 min. 42 sec.</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 16

*Learnability of Both Sites in Stage Two*

<table>
<thead>
<tr>
<th></th>
<th>Rutgers Web Site</th>
<th>Queens Web Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correctness</td>
<td>Speed</td>
</tr>
<tr>
<td>Rutgers Subjects</td>
<td>N_{R}=15</td>
<td>77% 2 min. 14 sec.</td>
</tr>
<tr>
<td>Queens Subjects</td>
<td>N_{Q}=15</td>
<td>67% 2 min. 18 sec.</td>
</tr>
</tbody>
</table>

As indicated in Table 15, both Rutgers and Queens subjects were able to complete tasks correctly and faster on the Queens site. However, this finding is not confirmed in stage two. Table 16 suggests that subjects feel more comfortable using their home institution's site than a new site.

5.5 Satisfaction

Satisfaction is a multi-dimensional construct. This study applies the Likert scale to measure its directions and intensity (1= easy to use/high satisfaction, 5=difficult to use/low satisfaction). Factors of satisfaction are carefully identified as shown in the model (see Figure 2). In addition to the Likert scales used after each task and in the post-test questionnaire, participants' comments were solicited. There were also open-ended questions.

The use of the Likert scale is known as an economical way of measuring user satisfaction (Pearson, 1977).

5.5.1 Overall Satisfaction

The scores for calculating overall satisfaction were gathered from the Likert scales after completing each task (see Appendix C). As indicated in Table 17, the overall
satisfaction ratings by all subjects are very close between the two sites in both stages one and two.

Table 17

*Satisfaction Rating by Question (1=easy to use/high satisfaction, 5=difficult to use/low satisfaction)*

<table>
<thead>
<tr>
<th>Task</th>
<th>Rutgers site Stage 1</th>
<th>Rutgers site Stage 2</th>
<th>Queens site Stage 1</th>
<th>Queens site Stage 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>Find a book</td>
<td>3.2</td>
<td>2.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Task 2</td>
<td>Find a journal</td>
<td>1.9</td>
<td>2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>Task 3</td>
<td>Use database</td>
<td>3.4</td>
<td>2.7</td>
<td>1.8</td>
</tr>
<tr>
<td>Task 4</td>
<td>Use database</td>
<td>2.2</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Task 5</td>
<td>Use database</td>
<td>2.8</td>
<td>2.5</td>
<td>3.2</td>
</tr>
<tr>
<td>Task 6</td>
<td>Find encyclopedia</td>
<td>2.7</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Task 7</td>
<td>Find e-book</td>
<td>4.1</td>
<td>3.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Task 8</td>
<td>Find information</td>
<td>2.2</td>
<td>1.9</td>
<td>2.1</td>
</tr>
<tr>
<td>Task 9</td>
<td>Find information</td>
<td>2.0</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>2.6</td>
<td>2.5</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Table 18 is a closer examination of the overall satisfaction of Rutgers and Queens subjects respectively. The Queens site provided higher satisfaction in stage one. In stage two, subjects were more satisfied with their own sites.

Table 18

*Satisfaction by Site (1=high satisfaction, 5=low satisfaction)*

<table>
<thead>
<tr>
<th></th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Rutgers subjects</td>
<td>2.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>2.7</td>
<td>2.7</td>
</tr>
</tbody>
</table>
The satisfaction ratings for task 1 (locate a book) on the Rutgers site were worse than the corresponding ratings on the Queens site in both stages. Thirteen (32%) of the subjects (R1-1, R1-9, R2-1, R2-5, R2-6, R2-7, R2-9, R2-10, R2-11, R2-13, Q1-2, Q1-4, and Q2-15) commented that a simple search by title on the Rutgers catalog resulted in too many hits when it was supposed to be a concise search. The Rutgers catalog provided two options for searching: “contains” and “begins with.” A title search for Gone with the Wind using “contains” results in 42 hits while the same search string using “begins with” results in 12 hits. The default is set to “contains” by the system. Most users did not realize the difference and did not change it. R2-7 said, “I typed Gone with the Wind without author’s name in the search for title and it wasn’t the first search result, which I thought it should have been.” R2-9 said, “I would have thought the title would be closer to the top of the list.” R2-10 said, “I don’t like it when a lot of unrelated things show up.” Q2-15 said, “The online catalog is the main problem.”

Figure 7 shows the Rutgers online catalog screen for an online catalog search.
R2-13 was not aware that the Rutgers advanced search allows the combination of author and title in one search.

R2-1 was impressed by the Queens site whose search results were more concise. R2-2 indicated that it was easy to identify items on the Queens site by format. R2-5 also declared that it was very easy to find books on the Queens online catalog.

The participants had more difficulties on task 7 (find an e-book) on the Rutgers site in both stages. It was hard to dig out e-books from the structure of the Rutgers site. This is also reflected in the results for effectiveness and efficiency. The Queens site promotes e-books on the first page of its Web site (see Figure 3).

For the Queens site, task 6 (find electronic encyclopedia) was the one with which all subjects felt the least satisfied.

Table 19 summarizes subjects' comments on overall satisfaction.
Table 19

Subjects’ Comments on Overall Satisfaction

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-1, R1-9, R2-1, R2-5, R2-6, R2-7, R2-9, R2-10, R2-11, R2-13, Q1-2, Q1-4, and Q2-15</td>
<td>A simple search by title on Rutgers catalog resulted in too many hits when it was supposed to be a concise search.</td>
<td></td>
</tr>
<tr>
<td>R2-1</td>
<td>Search results were more concise.</td>
<td></td>
</tr>
<tr>
<td>R2-2</td>
<td>It was easy to identify items by format.</td>
<td></td>
</tr>
<tr>
<td>R2-5</td>
<td>Easy to find books in Queens online catalog.</td>
<td></td>
</tr>
<tr>
<td>R2-13</td>
<td>Was not aware that Rutgers advanced search allows the combination of author and title in one search.</td>
<td></td>
</tr>
</tbody>
</table>

5.5.2 Ease of Use

Table 20 indicates the ratings of ease of use by Rutgers and Queens subjects in both stages one and two. The Queens site was rated easier to use than the Rutgers site by both Rutgers and Queens subjects in stage one. However, in stage two, subjects rated their own institution’s site easier to use. It is worth noting that the Rutgers site received a better rating in stage two, after the revamp of its Web site appearance.

Table 20

Ratings of Ease of Use (1=easy to use, 5=difficult to use)

<table>
<thead>
<tr>
<th>Ease of use</th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Rutgers subjects</td>
<td>2.6</td>
<td>2.5</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>3.0</td>
<td>2.7</td>
</tr>
</tbody>
</table>
a. Rutgers Site:

Q2-9 said the Rutgers site was easy to use: “Very easy to get around.” Q2-14 said, “It is set up well.” Q2-12 said he was new to the Rutgers site and could follow directions easily. Q2-13 said the left navigational bar made the site very easy to use.

R2-2 liked the new design of the Rutgers site. It made it easier to find articles.

R2-14 said the Rutgers site provided a good description of each section such as search online catalog, find articles, etc. These descriptions made it easy to use the site.

R2-8 said it was easy to use the Rutgers site for those tasks listed on the side bar. However, it was hard to find other information.

R2-15 said the Rutgers site was easy to use although when it came to database searching, he was clueless.

Q2-4 complimented Rutgers OPAC. He said it was easy to use.

R2-3 felt that it was too much hassle to use the Rutgers site to locate books, and the Queens site was “surprisingly easy to use.” R2-5 shared the same comment.

Q2-8 noted that there were two interfaces for the Rutgers online catalog. She said this could be confusing (See Figures 7 and 8).
Q2-8 said, “A lot of things aren’t obvious. What the texts say do [sic] not necessarily correlate to everything that you can find. It is not very user-friendly for someone who has not used it before.”

b. Queens Site:

R1-4 and R2-11 said that the Queens site was easier to use than the Rutgers site because its design was more clear-cut. Q2-3 said the Queens site was easier to use than the Rutgers site because the Queens site in his view was organized better. Q2-4 also said that the Queens site was easier to use.
Q2-2 said the Queens online catalog was easy to use. Q2-3 said “Generally, it is good except for a few things.” He had difficulties in locating an encyclopedia. “I don’t see where they have them available. I am not even sure where to look. It is not a periodical. It is not a database.” Q2-6 said, “It is easy to use. It is easy to get to where you want to go.”

Q1-5 liked the pull-down menu on the Queens site which facilitated navigation. R2-2, however, did not like the pull-down menu. The menu quickly disappeared when mouse moved. R2-15 did not like it either.

Neither did R2-15 like the pop-up menu which appeared to the right of each entry on the Queens home page (See Figure 4). He said, “They are kind of irritating. They don’t respond very well to the mouse.” Q2-1 agreed with this. He said that the Queens site was easy to use as long as those pop-ups worked. “They don’t stay.” Q2-5 shared that opinion.

Q1-2 commented that the Queens site was not intuitive enough. R2-14 said, “I searched and searched and do not know where to go.”

R2-8 was confused by the many catalogs in CUNY+. She also remarked that the Queens site was not easy to learn.

Table 21 presents the subjects’ comments in a summary form.
Table 21

**Subjects' Comments on Ease of Use**

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2-9, Q2-14, Q2-12</td>
<td>Easy to use, set up well, can follow directions easily</td>
<td></td>
</tr>
<tr>
<td>Q2-13</td>
<td>Left navigational pane is good</td>
<td></td>
</tr>
<tr>
<td>R2-2</td>
<td>It is easier to find articles after the new design</td>
<td></td>
</tr>
<tr>
<td>R2-14</td>
<td>Good description to each section</td>
<td></td>
</tr>
<tr>
<td>Q2-4</td>
<td>OPAC is easy to use</td>
<td></td>
</tr>
<tr>
<td>R2-3, R2-5</td>
<td>Surprisingly easy to use</td>
<td></td>
</tr>
<tr>
<td>Q2-8</td>
<td>There are two IRIS interfaces which are confusing</td>
<td></td>
</tr>
<tr>
<td>R1-4, R2-11, Q2-3, Q2-4, Q2-6</td>
<td>Easy to use, clear-cut, well organized</td>
<td></td>
</tr>
<tr>
<td>Q2-2</td>
<td>Easy to use except locating encyclopedia</td>
<td></td>
</tr>
<tr>
<td>Q1-5</td>
<td>The pull-down menu makes the site easy to navigate</td>
<td></td>
</tr>
<tr>
<td>R2-2, R2-15, Q2-1, Q2-5</td>
<td>The pull-down menu goes away when mouse moves</td>
<td></td>
</tr>
<tr>
<td>Q1-2, R2-14, R2-8</td>
<td>Not intuitive enough, not easy to learn</td>
<td></td>
</tr>
</tbody>
</table>

It can be synthesized that subjects evaluate “ease of use” from the perspectives of “easy to get around,” “can follow directions easily,” “easy navigation,” “clear description,” “intuitive,” and “user-friendly.”

5.5.3 Organization of Information

Table 22 indicates the ratings of organization of information. The Queens site received better ratings from both Rutgers and Queens subjects in stage one. However,
the Rutgers site, after the redesign of its Web site, received better ratings in stage two by both Rutgers and Queens subjects.

Table 22

Ratings of Organization of Information (1=clear, 5=unclear)

<table>
<thead>
<tr>
<th>Organization of Information</th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Rutgers subjects</td>
<td>3.0</td>
<td>1.9</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>2.6</td>
<td>2.1</td>
</tr>
</tbody>
</table>

a. Rutgers Site:

a.1 Stage One:

R1-1 declared that Google was better than both the Rutgers and the Queens sites and that she uses Google rather than library sites for her information needs whenever she can.

Q1-2 considered the Rutgers site's organization of information inconsistent. Two subjects (R1-5 and Q1-5) said that it was hard to find e-books and e-encyclopedias on the Rutgers site. Q1-1, however, complimented the Rutgers site for being more in-depth and detailed, yet clearly organized.

a.2 Stage Two:

R2-1 said that the organization of the Rutgers site was fair. Some things were organized well while some were not. For example, the instructions for inter-library loan, remote access, and hours of the library were easy to find.
R2-2 and Q2-4 said the organization of the Rutgers site was clear. Five subjects (R2-7, R2-8, R2-10, R2-15, and Q2-2) said Rutgers site was “well-organized.” R2-12 said “It is better organized” than the old version. R2-14 said, “It is easy to look up things.” Q2-5 said the Rutgers site’s organization was good: “Not much going on.” She also liked the left navigational pane which made it easy to get to common tasks.

Q2-13 also liked the placement of the navigational bar on the Rutgers site, which was placed on the left, while Queens placed it on the top. She said users look for information on the left or right. They don’t look on the top. She thinks the top is for the banner.

Q2-3 said, “I think it could be better. You don’t know from the front page where the electronic sources are. You have to go down to Reference. Why hide under Reference? I think they should tailor the Web site to more average user.” Q2-6 said the arrangement of electronic resources was unclear.

b. Queens Site:

R1-1 and R2-5 complimented the Queens site’s organization and layout. It was simple and straightforward. R2-2 said the Queens site was logical and straightforward. R2-3 thought the designers of the Queens site know what the users are looking for and organize it well. R2-10 said, “I think there are a lot of useful links. I think it is pretty well-organized.” R2-15 said, “For the most part, it is organized pretty well.” Q2-3 said, “I would say it is pretty good.” Q2-1 said the Queens site’s links were self-explanatory. Q2-8 said, “I think it is better than Rutgers site.”
R2-12 and R2-13 said the organization of the Queens site was unclear. Four subjects (R2-11, R2-12, R2-14, and R2-15) pointed out that the NetLibrary collections should have been listed with “e-books” instead of or in addition to “library catalogs” (see Figures 4 and 5). Q2-2 pointed out that “Under the current ‘e-books’ section, there are no e-book collections.” It was noticed on the October 18, 2004 experiment that the Queens site had corrected this oversight. The e-book icon was re-named “e-reference.” NetLibrary was listed under both “library catalogs” and “e-reference.” As of that date, twenty-six subjects of stage two had already completed their sessions and there were only four more to go.

R2-13 did not understand the distinction between “e-reference sources” and “periodicals research.”

Q2-15 said, “A lot of things are a little bit redundant. To me there is not much of a difference between e-Resources, e-Reference Sources, and Periodical Research. They could be all grouped together.”

R2-12 said, “There aren’t enough links on the home page. The things I am looking for are too many layers down.”

Q2-6 said, “I think it is organized poorly overall.”

Table 23 summarizes subjects’ comments on organization of information.
Table 23

Subjects’ Comments on Organization of Information

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2-2, R2-7, R2-8, R2-10, R2-12, R2-14, R2-15, Q1-1, Q2-2, Q2-4, Q2-5</td>
<td>Well organized</td>
<td></td>
</tr>
<tr>
<td>Q1-1</td>
<td>More in-depth and detailed than Queens site</td>
<td></td>
</tr>
<tr>
<td>Q2-5</td>
<td>Likes the left navigational pane which lists common tasks</td>
<td></td>
</tr>
<tr>
<td>Q2-13</td>
<td>Common tasks are listed in navigational pane on the left. This is better than Queens site.</td>
<td>Common tasks are grouped as pull-down menu on the top. The top is often looked on as banner. Rutgers’ arrangement is better.</td>
</tr>
<tr>
<td>Q1-2, R2-1</td>
<td>Organization inconsistent</td>
<td></td>
</tr>
<tr>
<td>R1-5, Q1-5</td>
<td>Hard to find e-books and e-encyclopedia</td>
<td></td>
</tr>
<tr>
<td>Q2-3, Q2-6</td>
<td>Hard to find electronic resources</td>
<td></td>
</tr>
<tr>
<td>R1-1, R2-2, R2-3, R2-5, R2-10, Q2-1, Q2-3, Q2-8</td>
<td>Organization and layout simple, logical, and straightforward</td>
<td></td>
</tr>
<tr>
<td>R2-12, R2-13</td>
<td>Organization unclear</td>
<td></td>
</tr>
<tr>
<td>R2-11, R2-12, R2-14, R2-15</td>
<td>NetLibrary should be grouped with e-books</td>
<td></td>
</tr>
<tr>
<td>R2-12</td>
<td>Not enough links on home page</td>
<td></td>
</tr>
<tr>
<td>Q2-6</td>
<td>Poorly organized</td>
<td></td>
</tr>
<tr>
<td>R2-13, Q2-15</td>
<td>Unclear distinction between e-resources, e-reference sources, and periodicals research</td>
<td></td>
</tr>
</tbody>
</table>

It can be synthesized that subjects evaluate “organization of information” from the perspectives of “simple,” “straightforward,” “logical,” “easy to look up things,” and “placing common tasks upfront.”
5.5.4 Terminology

Table 24 indicates how subjects rate the terminology used in both sites and the clarity of the categories' labeling. The Rutgers site received better ratings than the Queens site in both stages one and two and by both Rutgers and Queens subjects.

Table 24

*Ratings of Terminology (1=clear, 5=unclear)*

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Rutgers subjects</td>
<td>3.0</td>
<td>2.2</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>2.4</td>
<td>1.7</td>
</tr>
</tbody>
</table>

a. Rutgers Site:

Eight subjects expressed concerns about the terminology on the Rutgers site. R1-1 commented that the terminology used on the Rutgers site assumed that users have a common sense of library terms. R1-3 did not know the difference between index and database. R2-9 did not know what an encyclopedia is. Q1-5 and R2-9 thought that “articles,” “encyclopedia,” “serial,” “periodicals,” and “index” are library-oriented terms. The organization of the library’s Web site requires users to know the terminology before they can find out where to look for information. R2-9 said “It is tricky.” R2-7 felt the terminology used on the Rutgers site could be clearer. She would like to see explanations/definitions about articles, periodicals, databases, indexes, online catalogs, encyclopedia, etc. Q2-1 did not know what IRIS is. He suggested the addition of “catalog” next to IRIS. Q2-5 said “Connect from Home” is too vague. She thought it meant a connection to the catalog.
Six subjects complimented the Rutgers site’s terminology. R2-1 remarked that it was straightforward. R2-10 thought the terminology was “general, easy for people to understand.” R2-13 and Q2-5 felt that Rutgers has used a mix of both library terms and general user’s terms. Q2-9 said Rutgers “actually did it right.” Q2-15 said the terminology used on the Rutgers site was “pretty generic.”

b. Queens Site:

Five users expressed concerns about the terminology used on the Queens site. R2-14 said the Queens site did not label sections clearly. R2-13 felt the terminology used on the Queens site was confusing. He pointed out several problems such as “e-reference,” “Periodical Research,” and “Subject Guides and Websites.” Q2-5 said, “The terminology used in Queens site is still too librarian-oriented. [Users] don’t necessarily know the differences between electronic resources and Websites. I think that is confusing.” Q2-11 said the term “e-Resources” was somewhat vague. R2-12 said that users needed to be aware that they must look for books on the online catalog and look for articles through “Periodicals Research.” She said, “There should be explanations.”

Six subjects complimented the Queens’ site’s terminology. Q2-15 said, “It is good. It is fine. It is simple.” R2-2 and R2-10 said the Queens site labeled everything very clearly: “It is understandable.” Q2-1 said the Queens site’s terminology was clear and did not have much library jargon. Q2-9 said the terminology used on the Queens site was very straightforward: “Most people would be able to get it.” But he thought the labels were not helpful without tutorials or other help. He remarked that task 3 of the post-test questionnaire asked two separate questions, one about terminology and one
about the label. A section could be clearly labeled, but users still might not know what it means. Q2-13 said, “I think it is very clear. I think it is me.”

Table 25 summarizes the subjects’ comments on terminology.

Table 25

Subjects’ Comments on Terminology

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2-1, R2-10, R2-13, Q2-5,</td>
<td>Terminology is straightforward, easy to understand</td>
<td></td>
</tr>
<tr>
<td>Q2-9, Q2-15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1-1, R1-3, R2-7, R2-9,</td>
<td>Terminology is not user-oriented</td>
<td></td>
</tr>
<tr>
<td>Q1-5, Q2-1, Q2-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2-2, R2-10, Q2-1, Q2-9,</td>
<td>Clear labeling, terminology</td>
<td></td>
</tr>
<tr>
<td>Q2-13, Q2-15</td>
<td>simple and understandable</td>
<td></td>
</tr>
<tr>
<td>R2-12, R2-13, R2-14, Q2-5,</td>
<td>Sections not labeled clearly, terminology</td>
<td></td>
</tr>
<tr>
<td>Q2-11,</td>
<td>confusing</td>
<td></td>
</tr>
</tbody>
</table>

It can be synthesized that subjects evaluate “terminology” from the perspectives of “simple,” “straightforward,” “understandable,” “generic,” “label sections clearly,” “no jargon,” “clear descriptions/explanations,” and “from user’s perspective.”

5.5.5 Attractiveness

Table 26 indicates how subjects rate the degree of visual attractiveness of both sites. It is interesting to find that subjects give their home institution’s site better rating, in both stages one and two.
Table 26

*Ratings of Attractiveness (1=attractive, 5=unattractive)*

<table>
<thead>
<tr>
<th>Attractiveness</th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Rutgers subjects</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>3.0</td>
<td>2.8</td>
</tr>
</tbody>
</table>

a. Rutgers Site:

a. 1 Stage One:

Four subjects (R1-2, Q1-2, Q1-3, and Q1-5) commented that the Rutgers site should use more graphics. Q1-4 suggested the use of more space between the entries of Rutgers IRIS search results.

a.2 Stage Two:

R2-2 thought the Rutgers site was attractive. R2-9 said: “It is readable. You don’t want too much going on graphically when the site is basically about words. It is readable. This is important.”

R2-1 commented that she liked the old Rutgers site better; she considered the current one gloomy. R2-12 and Q2-14 disliked the grey on the left navigational bar. R2-12 said, “It is too dark.” Q2-14 said, “It doesn’t stand out.” R2-10, however, thought the Rutgers site was more attractive than the Queens site in terms of color. R2-15 liked the red color on the Rutgers site. He said, “Red is easy on the eyes.” Q2-9 said the Rutgers site was more visually attractive than the Queens site in that it was not always red. “Even the red has got some colors in it so that it won’t be a problem for red-blind users.”
R2-5 felt the Rutgers site was too complicated.

R2-12 and Q2-2 said the font used on the Rutgers site was too small. Q2-2 also would like to see more graphics on the Rutgers site.

R2-13 thought the current version was more attractive than the old one. The old version had too much information cramped on the screen.

Q2-6 said there were too many words and paragraphs. “It seems like you have to read a lot to find things.”

Q2-10 said the Rutgers site was “boring, not visually interesting.”

b. Queens Site:

R2-1 liked the color of the Queens site. Q2-1 also liked it. R2-7 thought the Queens site visually appealing. R2-15 said it was a nice-looking site, and R2-12 liked the large font: “It is easy to read.” Q2-13 said, “I think it is clean, not too busy.” Q2-15 said, “Great colors.”

Q2-9 said the home page was attractive, simple, clear, and straightforward. However, he said, “I don’t like red.” He said that red could be a problem for visually-impaired users.

Q2-2 said the Queens site was more attractive than the Rutgers site. The Queens Library Web site had a large picture of the library and placed important functions such as catalogs, periodicals, etc. so as to dominate the home page.

R2-1 would like to see bigger lettering on the Queens site.
R2-5 thought the Queens site was too simple. R2-8 said it was too dark. R2-10 said, “I think it could be made better. It is boring. There should be more graphics and images.”

Q2-7 said the Queens site was “too crowded.”

Table 27 summarizes the subjects’ comments on visual attractiveness.

**Table 27**

*Subjects’ Comments on Visual Attractiveness*

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-2, Q1-2, Q1-3, Q1-5, Q2-2</td>
<td>Should use more graphics</td>
<td></td>
</tr>
<tr>
<td>Q1-4</td>
<td>Should use more space between entries</td>
<td></td>
</tr>
<tr>
<td>R2-2, R2-10, R2-15</td>
<td>Attractive, readable</td>
<td></td>
</tr>
<tr>
<td>R2-1, R2-12, Q2-14</td>
<td>Gloomy, too dark</td>
<td></td>
</tr>
<tr>
<td>R2-5, Q2-6</td>
<td>Too complicated</td>
<td></td>
</tr>
<tr>
<td>R2-12, Q2-2</td>
<td>Font too small</td>
<td></td>
</tr>
<tr>
<td>Q2-10</td>
<td>Boring</td>
<td></td>
</tr>
<tr>
<td>R2-1, R2-7, R2-15, Q2-1, Q2-2, Q2-9</td>
<td>Likes the color, visually appealing</td>
<td></td>
</tr>
<tr>
<td>R2-12</td>
<td>Likes the large font, easy to read</td>
<td></td>
</tr>
<tr>
<td>Q2-13</td>
<td>Clean, not too busy</td>
<td></td>
</tr>
<tr>
<td>R2-1</td>
<td>Need larger lettering</td>
<td></td>
</tr>
<tr>
<td>R2-5</td>
<td>Too simple</td>
<td></td>
</tr>
<tr>
<td>R2-8</td>
<td>Too dark</td>
<td></td>
</tr>
<tr>
<td>R2-10</td>
<td>Should have more graphics and images</td>
<td></td>
</tr>
<tr>
<td>Q2-7</td>
<td>Over crowded</td>
<td></td>
</tr>
</tbody>
</table>

It can be synthesized that subjects evaluate “attractiveness” from the perspectives of “appropriate graphics,” “readability,” “appropriate color,” “not too complicated,” and “appropriate size of font.”
5.5.6 Mistake Recovery

Table 28 displays the subjects’ ratings of both sites’ mistake recovery features.

The Rutgers site received a better rating from both Rutgers and Queens subjects in stage one. However, in stage two, each subject rated his/her own site better at mistake recovery.

Table 28

*Ratings of Mistake Recovery (1=easy, 5=difficult)*

<table>
<thead>
<tr>
<th>Mistake Recovery</th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Rutgers subjects</td>
<td>2.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>2.0</td>
<td>2.4</td>
</tr>
</tbody>
</table>

a. Rutgers Site:

a.1 Stage One:

Five subjects (R1-2, R1-3, R1-4, Q1-2 and Q1-5) remarked that the Rutgers site had a better navigation bar on the top of its Web site, which made it easier to start over.

a.2 Stage Two:

Eight subjects complimented the Rutgers site on making it easy to recover from mistakes. R2-1, R2-2, R2-14, and Q2-2 declared that it is easy to start over at the Rutgers site. R2-2 said, “I can always find my way back.” R2-12 said it was very easy to get back to the Rutgers Libraries home page. R2-7 said that the new Rutgers site made it a lot easier to recover from mistakes than the old one did. R2-9 and Q2-6 noted that the “Back” button on the Rutgers site was useful.
Q2-5 said she was “stranded in electronic resources.”

b. Queens Site:

Five subjects complained about the Queens site’s mistake recovery feature. R2-1 and R2-14 found it hard to recover from mistakes on the Queens site. R2-5, R2-13, and Q2-12 said the Queens site made it hard to go back.

There were, however, three subjects (R2-4, R2-10, and Q2-2) who felt that the Queens site made it easy to go back. Q2-5 said, “It is okay overall.”

R2-14 and Q2-12 recommended a better “Help” section for the Queens site.

Table 29 summarizes the subjects’ comments on mistake recovery.

Table 29
Subjects Comments on Mistake Recovery

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-2, R1-3, R1-4, R2-1, R2-2, R2-7, R2-9, R2-12, R2-14, Q1-2, Q1-5, Q2-2, Q2-6</td>
<td>Easy to start over by using the navigation bar on top of the page or the Back button</td>
<td></td>
</tr>
<tr>
<td>R2-4, R2-10, Q2-2</td>
<td>Easy to start over</td>
<td></td>
</tr>
<tr>
<td>R2-1, R2-5, R2-13, R2-14, Q2-12</td>
<td>Difficult to recover from mistakes</td>
<td></td>
</tr>
<tr>
<td>R2-14, Q2-12</td>
<td>Need better Help section</td>
<td></td>
</tr>
</tbody>
</table>

It can be synthesized that subjects evaluate “mistake recovery” from the perspective of “easy navigation.” It is helpful to have a navigation bar and a Back button. It is also important to be able to start over. A “Help” section is suggested.
5.5.7 Overall Reaction

The scores for calculating overall reaction were from the Likert scales in the post-test questionnaires (see Appendix D). This gave participants another opportunity to rate the test sites and to provide their overall reactions after examining specific areas such as ease of use, organization of information, terminology, attractiveness, and mistake recovery. Table 30 displays the overall reaction to both the Rutgers and the Queens sites. It turns out that subjects were more pleased with the Queens site in stage one, but more pleased with the Rutgers site in stage two, after Rutgers revamped its site.

Table 30

Ratings of Overall Reaction (1=satisfied, 5=unsatisfied)

<table>
<thead>
<tr>
<th>Overall Reaction</th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Rutgers subjects</td>
<td>2.4</td>
<td>2.1</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>3.2</td>
<td>2.2</td>
</tr>
</tbody>
</table>

a. Rutgers Site:

a.1 Stage One:

Four subjects (R1-2 and Q1-2, Q1-3, and Q1-5) declared that the overall design of the Rutgers site in stage one was too dry and simple. There should be more and better graphics.

a.2 Stage Two:

R2-2 said, “Rutgers site is helpful.” R2-15 said, “It is a good system. There is a lot of information.” Q2-5 said, “It is quite good. I am pretty satisfied with it.” Q2-6 said, “Overall I am satisfied.” Q2-9 said, “The site seems to be laid out very well.”

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
R2-10 said, "There is a lot of information. It is hard to digest it all. I think it is hard to find online article and things like that."

b. Queens Site:

R2-2 said that she is a satisfied customer of the Queens site.

R2-4 does not like the layout of the Queens site.

Table 31 summarizes the subjects' comments on overall reaction.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-2, Q1-2, Q1-3, Q1-5</td>
<td>Too dry and simple. Should use more graphics</td>
<td></td>
</tr>
<tr>
<td>R2-2, R2-15, Q2-5, Q2-6</td>
<td>Helpful. Good system. Has lots of information.</td>
<td></td>
</tr>
<tr>
<td>Q2-9</td>
<td>Laid out well</td>
<td></td>
</tr>
<tr>
<td>R2-10</td>
<td>Difficult to digest all information on the site.</td>
<td>Difficult to locate online journals.</td>
</tr>
<tr>
<td>R2-2</td>
<td></td>
<td>Satisfied.</td>
</tr>
<tr>
<td>R2-4</td>
<td></td>
<td>Dislike the layout.</td>
</tr>
</tbody>
</table>

5.6 Best Features

a. Rutgers Site:

a.1 Stage One:

R1-1 complimented the "Ask a librarian" feature. She has used it often.
Two subjects (R1-3 and R1-4) liked the “How do I?” section.

Q1-4 complimented the Rutgers site for the resources it contained. It was rich.

The same subject also complimented the News and Events section.

a.2 Stage Two:

Ten subjects (R2-2, R2-5, R2-9, R2-12, Q2-1, Q2-3, Q2-5, Q2-9, Q2-12, Q2-13) liked the left navigational pane which was new after Rutgers University Libraries redesigned its Web site. The left pane contained icons that lead you to common tasks (See Figure 9), including using online catalog, find articles, how do I, inter-library loan, and connect from home. R2-12 said “It is an excellent idea. Even if I were new to the site, it would be fairly easy.” Q2-3 said he preferred to have such items listed on the left as the Rutgers site does. The Queens site placed them on the top. He said, “Up there is a little hidden.” Q2-9 said the Rutgers site’s layout was very good.
Q2-2 praised the “Find Articles” option on the left navigational pane which linked to useful information on finding journal articles. Q2-5 complimented the Rutgers site on providing brief descriptions of each task on the left navigational pane and the fact that the one for “Find Articles” is especially good.

R2-3 and Q2-6 liked the “Search Website” feature which is an internal search engine to locate information within Rutgers University Libraries Web site.

R2-7 liked the Reserve function.

R2-8 liked the advanced search function on the Rutgers online catalog.

R2-11 liked the online inter-library request feature.
R2-15 and Q2-15 liked the subject guides.

Q2-7 said the best feature of the Rutgers site was the “easiness of finding what you want to know.”

Q2-14 said the Rutgers site’s best feature was “Ask the Librarian.”

Q2-15 said, “It’s a huge site. The site index is great.”

b. Queens Site:

R1-3 thought the Queens site index was great.

Q1-2 liked the links to other New York City library catalogs. R2-2, Q2-2, and Q2-11 complimented the integrated online catalog, CUNY+. R2-10 liked the ability to access a number of different libraries’ online catalogs from the Queens home page.

Q1-4 praised the content of the Queens site. It is rich, powerful, and easy to use.

Q2-5 said, "You can get to quite a lot of resources."

Q2-1 admired the drop-down menu which facilitated navigation.

R2-1 liked the search engine of the Queens site. Q2-9 shared that opinion.

R2-4 thought the Queens site made it easy to locate e-books.

R2-15, Q2-1, and Q2-14 liked the subject lists of databases. They helped find articles.

Q2-12 liked “Article Database, A-Z” in which all databases were listed alphabetically.

R2-15 also liked the way the Queens site displayed search results of the online catalog, which included author, title, year, format, and holdings, all in one line.

Q2-3 said the best feature of the Queens site is its “How do I” link.
Table 32 summarizes the subjects’ comments on best features.

**Table 32.**

**Subjects Comments on Best Features**

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-1, Q2-14</td>
<td>Ask a Librarian</td>
<td></td>
</tr>
<tr>
<td>R1-3, R1-4</td>
<td>How Do I?</td>
<td></td>
</tr>
<tr>
<td>Q1-4</td>
<td>News and Events section and the rich resources on Rutgers site</td>
<td></td>
</tr>
<tr>
<td>R2-2, R2-5, R2-9, R2-12, Q2-1, Q2-3, Q2-5, Q2-9, Q2-12, Q2-13</td>
<td>The left navigational pane</td>
<td></td>
</tr>
<tr>
<td>Q2-2, Q2-5</td>
<td>Find Article</td>
<td></td>
</tr>
<tr>
<td>R2-3, Q2-6</td>
<td>Search Website</td>
<td></td>
</tr>
<tr>
<td>R2-7</td>
<td>Reserve</td>
<td></td>
</tr>
<tr>
<td>R2-8</td>
<td>Advanced Search</td>
<td></td>
</tr>
<tr>
<td>R2-11</td>
<td>Inter-library Loan</td>
<td></td>
</tr>
<tr>
<td>R2-15, Q2-15</td>
<td>Subject Guides</td>
<td></td>
</tr>
<tr>
<td>Q2-15</td>
<td>Site Index</td>
<td></td>
</tr>
<tr>
<td>R1-3</td>
<td>Site Index</td>
<td></td>
</tr>
<tr>
<td>Q1-2, R2-10</td>
<td>Links to other catalogs</td>
<td></td>
</tr>
<tr>
<td>R2-2, Q2-2, Q2-11</td>
<td>CUNY+</td>
<td></td>
</tr>
<tr>
<td>Q1-4</td>
<td>Content</td>
<td></td>
</tr>
<tr>
<td>Q1-5</td>
<td>Drop down menu</td>
<td></td>
</tr>
<tr>
<td>R2-1, Q2-9</td>
<td>Search Engine</td>
<td></td>
</tr>
<tr>
<td>R2-4</td>
<td>Easy to locate e-books</td>
<td></td>
</tr>
<tr>
<td>R2-15, Q2-1, Q2-14</td>
<td>Subjects lists of databases</td>
<td></td>
</tr>
<tr>
<td>Q2-12</td>
<td>Alphabetical list of databases</td>
<td></td>
</tr>
<tr>
<td>R2-15</td>
<td>Display of online catalog search results</td>
<td></td>
</tr>
<tr>
<td>Q2-3</td>
<td>How Do I?</td>
<td></td>
</tr>
</tbody>
</table>

**5.7  Worst Features**

a. Rutgers Site:

Several subjects (R1-2, R2-2, R2-7, R2-9) noted that users needed to have a good sense of library terminology or they would be lost in the clutter of information on the Rutgers site. For example, how would one find articles in an encyclopedia?
*Encyclopedia* was in *reference sources*, but not in the *databases*. This may be obvious to librarians, but could be confusing to novice users. R2-7 made similar comments about the Queens site.

Q1-4 complained that there was little relationship between entries and pop-ups.

Three subjects (R2-1, R2-7, and R2-8) said it is difficult to find articles on the Rutgers site. Three subjects (R2-3, R2-8, and R2-12) did not understand the electronic databases on the Rutgers site and would have liked to receive more instruction. R2-9 declared that she had difficulty in combining terms when she searched databases for task 3.

R2-7 commented that it was hard to locate e-books on the Rutgers site.

Thirteen subjects (R1-1, R1-9, R2-1, R2-5, R2-6, R2-7, R2-9, R2-10, R2-11, R2-13, Q1-2, Q1-4, and Q2-15) expressed dissatisfaction with the Rutgers online catalog, IRIS, because it provided too many hits when *contains* was used. For example, R2-5 said, “I am confused that there are many books displayed on the screen. It makes me to [sic] spend more time.” R2-3 and Q2-2 requested better instructions on how to locate books efficiently on the Rutgers online catalog. Q2-15 said, “The online catalog is the main problem. I would use the Queen site as a model.”

R2-15 did not like the way search results of the online catalog were displayed.

Q2-1 said the worst feature was that it was hard to get back to the Rutgers home page from an affiliated database page. Q2-5 shared that opinion. She said she was “stranded” in e-resources.
b. Queens Site:

Six subjects (R1-2, R1-3, R2-1, R2-12, Q1-1, and Q2-9) noted that the Queens site’s weakness was its navigation. It was hard to go back to the Queens Library home page.

R1-4 suggested the use of more contrast colors in the Queens navigation bar. The current colors were too dark and it was hard to see the words in those buttons.

Q1-2 thought that the Queens Library oversold e-books in its Web site.

Q1-4 and Q2-12 suggested the addition of contact information (phone number, e-mail address, etc.) on the home page.

R2-2 thought that the periodical record display on the Queens site made it difficult to tell if the library had a current subscription to the journal. There should be clearer holdings information.

R2-4 thought the worst feature of the Queens site was the way to search for an encyclopedia.

The Queens site provided commands for the next lower level using drop-down menu and pop-ups. R2-9 and Q2-12 complained that the Queens site’s drop-down menu did not stay. It disappeared when the mouse moved. Q2-1 said the worst feature of the Queens site was that the pop-ups did not stay. Q2-4 also pointed out these problems. He said, “They are too sensitive.”

R2-13 and Q2-5 felt the worst feature of the Queens site was the confusing terminology on its home page, such as e-reference sources, periodicals research, subject guides and Websites. Q2-15 shared that opinion, commenting that the Queens site used too much similar-sounding terminology, such as e-Resources, e-Reference, e-Reference...
Sources, Periodicals Research, and e-Periodicals. He said, “Confusing to me.” Q2-2 and Q2-5 said the Queens site used too much “E-” in its labels, such as e-forms, e-Reserve, e-periodicals, e-Reference Sources, etc.

Q2-13 said there was too much information cramped on the top. She suggested the re-arrangement of the sequence of the top icons to list QC Home first, then CUNY Libraries.

Q2-14 said the worst feature of the Queens site was the difficulty in connecting from home. The proxy server was not reliable.

Table 33 summarizes the subjects’ comments on worst features.
### Table 33.  
\textit{Subjects' Comments on Worst Features}

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-2, R2-2, R2-7, R2-9</td>
<td>Terminology and organization</td>
<td></td>
</tr>
<tr>
<td>R2-7, R2-13, Q2-2, Q2-5, Q2-15</td>
<td></td>
<td>Terminology and organization</td>
</tr>
<tr>
<td>Q1-4</td>
<td>Little relationship between entries and pop-ups</td>
<td></td>
</tr>
<tr>
<td>R2-1, R2-7, R2-8</td>
<td>Difficult to find articles</td>
<td></td>
</tr>
<tr>
<td>R2-3, R2-8, R2-12</td>
<td>Descriptions of databases</td>
<td></td>
</tr>
<tr>
<td>R2-7</td>
<td>Difficult to find e-books</td>
<td></td>
</tr>
<tr>
<td>R1-1, R1-9, R2-1, R2-5, R2-6, R2-7, R2-9, R2-10, R2-11, R2-13, Q1-2, Q1-4, and Q2-15</td>
<td>Too many hits for a simple search in online catalog. The default is set as “contains” and not “begins with”</td>
<td></td>
</tr>
<tr>
<td>R2-15</td>
<td>The display of online catalog search results</td>
<td></td>
</tr>
<tr>
<td>Q2-1, Q2-5</td>
<td>Difficult to get back to library home page from database pages</td>
<td></td>
</tr>
<tr>
<td>R1-2, R1-3, R2-1, R2-12, Q1-1, Q2-9</td>
<td>Difficult to get back to library home page</td>
<td></td>
</tr>
<tr>
<td>R1-4</td>
<td>Color in navigation bar</td>
<td></td>
</tr>
<tr>
<td>Q1-2</td>
<td>e-books are oversold in home page</td>
<td></td>
</tr>
<tr>
<td>Q1-4, Q2-12</td>
<td>Lack of contact information on home page</td>
<td></td>
</tr>
<tr>
<td>R2-2</td>
<td>Holdings information</td>
<td></td>
</tr>
<tr>
<td>R2-4</td>
<td>Encyclopedia searching</td>
<td></td>
</tr>
<tr>
<td>R2-9, Q2-1, Q2-4, Q2-12</td>
<td>Drop-down commands do not stay</td>
<td></td>
</tr>
<tr>
<td>Q2-13</td>
<td>Too much information on top of the home page</td>
<td></td>
</tr>
<tr>
<td>Q2-14</td>
<td>Difficult to connect from home</td>
<td></td>
</tr>
</tbody>
</table>
5.8 Desired Features

a. Rutgers Site:

R1-1 suggested the addition of “history of searching” as a reminder of previous attempts.

The e-book collection at Rutgers was hidden in “Indexes and Databases” in stage one. R1-2 and R1-4 suggested that e-books be promoted to the Library’s home page, in the manner of the Queens site (see Figure 3). R2-9 suggested providing more obvious links to e-books and encyclopedias. R2-13 suggested that different formats of the same materials should be in the same cataloging record or at least linked. For example, he said there should be a link between the print and e-book versions of *The Story of Mankind*. Q2-1 would like to have clear instructions on how to locate e-books in the Rutgers online catalog.

R1-5 and R2-13 expressed the need for federated searching (or meta-searching) across all databases. The different interfaces of databases were confusing.

R2-8 would like to see the Rutgers online catalog work like Google, which uses natural language searching.

R2-11 would like the online catalog search results to relate more to the search string. The search results for books should display the edition statement.

R2-12 wanted better ways for users to choose electronic databases. R2-15 suggested more detailed guides on how to use each specific database.

R2-13 would have liked to have a personal portal in which his frequently performed functions were conveniently listed for him.

R2-14 suggested having a section labeled “Encyclopedia.”
Q2-3 thought that all Rutgers electronic resources should be listed together. Databases were currently listed under *Find Articles* while the other electronic sources such as e-books and encyclopedias were listed under *Electronic Reference Sources*. He said, “It is confusing.”

Q2-8 suggested more “How do I” options.

Q2-9 said the “Rutgers Libraries” button should appear on every page so that users can always start over from the home page.

b. Queens Site:

R1-4 recommended the inclusion of a prominent “Frequently Asked Questions” or “How do I?” column. Q2-12 asked for a better “Help” menu to assist navigation.

Q1-4 suggested a “New Announcement” column to bring users up-to-date on new services, features, or activities.

Q1-5 suggested better highlighting for inter-library loan and the placement of holds on the pull-down menu.

R2-12 recommended the addition of “News” section.

R2-2 suggested “History of searching.” This would allow the ability to review and re-use previous searching strings in a session. The history could be wiped out when the user logs off.

R2-8 noticed that the Queens site provided *phrase search* which connected related terms with “or” inside a box, that is, “aged or elderly”. She said this was new to her.
R2-8 also commented on the need for advanced search on the Queens online catalog. R2-10 did not understand what guided search meant, and thought that explanations should be included.

R2-13 would have liked to have “New Search” in the Queens online catalog to start a search over. He also suggested that periodical title be separated from title searching. In addition, he suggested that search results be sorted by format.

Q2-2 said, “Information listed under too many choices. Everything should be accessible through catalogs.”

Q2-5 would have liked to see more explanations about existing features.

Q2-6 asked for more explanations on electronic resources.

Q2-7 would have liked to be able to specify the Queens College collections in CUNY+ and stay there, instead of being required to specify it every time.

Table 34 summarizes the subjects’ comments on desired features.
Table 34.

**Subjects’ Comments on Desired Features**

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-1</td>
<td>History of Searching</td>
<td></td>
</tr>
<tr>
<td>R1-2, R1-4, R2-9, Q2-1</td>
<td>Promotion of e-books</td>
<td></td>
</tr>
<tr>
<td>R2-9, R2-14</td>
<td>Promotion of encyclopedia</td>
<td></td>
</tr>
<tr>
<td>R2-13</td>
<td>Linking of different formats of materials in online catalog</td>
<td></td>
</tr>
<tr>
<td>R1-5, R2-13</td>
<td>Federated searching</td>
<td></td>
</tr>
<tr>
<td>R2-8</td>
<td>Natural language searching</td>
<td></td>
</tr>
<tr>
<td>R2-11</td>
<td>Display of edition statement on online catalog search results</td>
<td></td>
</tr>
<tr>
<td>R2-12, R2-15</td>
<td>Help in selecting databases</td>
<td>Descriptions on databases</td>
</tr>
<tr>
<td>Q2-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2-13</td>
<td>Personal portal</td>
<td></td>
</tr>
<tr>
<td>Q2-3</td>
<td>Re-arrangement of all e-resources</td>
<td></td>
</tr>
<tr>
<td>Q2-8</td>
<td>More How do I?</td>
<td></td>
</tr>
<tr>
<td>Q2-9</td>
<td>Every page should have an icon to start over from home page</td>
<td></td>
</tr>
<tr>
<td>R1-4</td>
<td></td>
<td>Frequently Asked Questions</td>
</tr>
<tr>
<td>R1-4, Q2-5</td>
<td></td>
<td>How Do I?</td>
</tr>
<tr>
<td>R2-12</td>
<td>Help</td>
<td></td>
</tr>
<tr>
<td>Q1-4, R2-12</td>
<td>New Announcement</td>
<td></td>
</tr>
<tr>
<td>Q1-5</td>
<td>Highlight Inter-Library Loan and Hold</td>
<td></td>
</tr>
<tr>
<td>R2-2</td>
<td>History of Searching</td>
<td></td>
</tr>
<tr>
<td>R2-8</td>
<td>Advanced Search</td>
<td></td>
</tr>
<tr>
<td>R2-10</td>
<td>Explanation of Guided Search</td>
<td></td>
</tr>
<tr>
<td>R2-13</td>
<td>New Search function in online catalog</td>
<td></td>
</tr>
<tr>
<td>R2-13</td>
<td>Separate periodical search from title search</td>
<td></td>
</tr>
<tr>
<td>R2-13</td>
<td>Sort online catalog search results by format</td>
<td></td>
</tr>
<tr>
<td>Q2-2</td>
<td>Let online catalog be the gateway to all resources</td>
<td></td>
</tr>
<tr>
<td>Q2-7</td>
<td>Limit search to Queens College only</td>
<td></td>
</tr>
</tbody>
</table>
5.9 Lostness

The study examined the user lostness issue in stage two. It found that fourteen subjects (46%) felt lost on the Rutgers site while seventeen subjects (57%) felt lost on the Queens site (see Table 35). An examination of Rutgers and Queens subjects separately indicated that subjects were more frequently lost on a new site than on their institution’s site.

Table 35

User Lostness (N2=30)

<table>
<thead>
<tr>
<th></th>
<th>Rutgers Site</th>
<th>Queens Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rutgers Subjects</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Queens Subjects</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>17</td>
</tr>
</tbody>
</table>

a. Rutgers Site:

R2-1 said she felt lost not because of the Rutgers site’s design, but because she lacked confidence in performing the tasks. It had more to do with the levels of difficulty of the tasks. Q2-1 felt the same way. He said the IRIS catalog was very straightforward, but there were no directions on how to search e-books and encyclopedias.

Q2-12 said he was somewhat lost in the beginning but things quickly make sense.

Some subjects did not feel lost. For instance, Q2-9 said, “I can figure out exactly what I want to look for.”

b. Queens Site:

R2-1 was lost on the Queens site because the navigation was not specific enough.

She did not know where she was going.
R2-2 was lost on the Queens site because some windows disabled the “Back” feature, which made navigation cumbersome. R2-15 said he was lost because he couldn’t go back to the Queens library’s home page.

Q2-5 said she occasionally felt frustrated but not lost.

Q2-9 sometimes felt lost because he did not have a sense of being able to find what he was looking for. For instance, he felt lost when he was trying to find encyclopedia on the Queens site. “There should be a ready reference section on the site, including encyclopedia and dictionary. It is probably there some place, but I couldn’t find it.”

Q2-10 felt lost when she performed database searching.

5.10 Navigation

The navigation issue was examined in stage two of the present study. Twenty-four subjects (71%) said the Rutgers site was easy to navigate, while nineteen subjects (56%) said the Queens site was easy to navigate. Table 36 indicates further breakdown by Rutgers and Queens subjects.

Table 36.

Ease of Navigation ($N_2=30$)

<table>
<thead>
<tr>
<th></th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rutgers subjects</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>19</td>
</tr>
</tbody>
</table>
a. Rutgers Site:

R2-4 felt the new design of the Rutgers site made it easier to navigate. Q2-1 said the links on the Rutgers site stayed in place and were self-explanatory. Q2-5 said the Rutgers site was easy to navigate.

Q2-10 had a problem going back to the home page.

b. Queens Site:

Q2-1 thought the Queens site provided good linkage. However, the links on the drop-down menu disappeared easily when the mouse moved. Q2-15 said the links were “incredible.”

R1-3 said the Queens site was difficult to navigate. R2-10 said, “Comparing to Rutgers site, I think it is a little bit more difficult to navigate.” In addition, there were six subjects (R1-2, R1-3, R2-1, R2-12, R2-15 and Q1-1) who noted that it was hard to go back to the Queens Library home page to recover from mistakes. R2-15 said, “It is easy to navigate from home page but hard to get back.” Q2-9 asked for more consistency on the Queens navigation bar: “Make them the same place all the time.”

R2-13 and R2-15 suggested that the Queens site put a “New Search” option in its online catalog.

5.11 Click Cost

The click cost issue was first examined in stage two of the present study. Forty-four subjects (73%) declared that they expected the click(s) to lead them eventually to the correct answer. Table 37 is a breakdown by Rutgers and Queens subjects.
Table 37.

*Click Cost (N<sub>2</sub>=30)*

<table>
<thead>
<tr>
<th></th>
<th>Rutgers site</th>
<th>Queens site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rutgers subjects</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Queens subjects</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>22</td>
</tr>
</tbody>
</table>

R2-2 and Q2-9 said they did not expect a click to answer the question immediately, but they did expect it to give them information relevant to the link. R2-3 said a click was supposed to imply what you could expect. R2-15 said, "I expect the click will take me to what it says it will." Q2-2 said, "I expect, but it doesn’t always.” R2-4 and R2-13 said that they normally would expect a click to lead them further along the right track to the correct answer. However, neither subject had this expectation on the Queens site. They did answer Yes on the Rutgers site. R2-9 said she was not really sure what would happen when she clicked on the Rutgers site. She said that she expected to get where she needed to be from the Queens main page, but she was less confident within electronic databases. R2-10 said she never expected to find information at the first try on the Rutgers site, but she marked Yes for the Queens site. R2-11 said it pretty much depended on the tasks she performed. For example, she did not have this expectation when she was doing database searching for periodical articles, but she had it when she did simple bibliographic searching on an online catalog. R2-12 said she just hoped for the best. Q2-6 said about the Queens site, “I hoped, but it didn’t.” Q2-13 said, “Of course, who doesn’t?”
5.12 Relationships among Effectiveness, Efficiency, and Satisfaction

One of the research goals of the present study is to examine the relationships among effectiveness, efficiency, and satisfaction. This section reports the results of stages one and two respectively. The satisfaction scores used in this section were gathered when subjects completed each task (see Appendix C). The unit of analysis is search.

5.12.1 Effectiveness and Satisfaction

Effectiveness in this study was measured by whether the subjects answered the questions correctly. Two one-way analyses of variance were conducted to evaluate the relationships between correctness of answers and satisfaction, for stages one and two respectively. The dependent variable was satisfaction (1=easy to use/high satisfaction to 5=difficult to use/low satisfaction) and the independent variable was correctness of answers (0=wrong answer and 1=correct answer).

The ANOVA for stage one was significant, $F(1, 163) = 57.57, p < .001$. The strength of the relationship between satisfaction and correctness of answers, as assessed by $\eta^2$, was strong, with the correctness of answer factor accounting for 26% of the variance of the dependent variable. Post-hoc tests were not performed because there were fewer than three levels.

The ANOVA for stage two was also significant, $F(1, 445) = 185.48, p < .001$. Because the $p$-value was less than .05, the null hypothesis that there were no differences between the groups was rejected. The $\eta^2$ of .29 indicates a strong relationship between
satisfaction and correctness of answers. Post-hoc tests were not performed because there were fewer than three levels.

Based on the findings of the two ANOVA analyses that there were significant differences in the means of satisfaction between right and wrong answers, an interpretation may be made that subjects felt less satisfied with the system when they failed to perform the task correctly.

Table 38 indicates the means of satisfaction ranking with regard to correctness of answers in both stages. The scores for both stages are close.

Table 38

<table>
<thead>
<tr>
<th></th>
<th>Satisfaction (Stage 1)</th>
<th>Satisfaction (Stage 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctness of Answer</td>
<td>N=11</td>
<td>N=30</td>
</tr>
<tr>
<td>Wrong Answer</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>1.17</td>
</tr>
<tr>
<td>Correct Answer</td>
<td>2.20</td>
<td>1.16</td>
</tr>
</tbody>
</table>

**5.12.2 Efficiency and Satisfaction**

Correlation coefficients were computed among time spent on performing tasks, number of steps, and satisfaction. Using the Bonferroni approach to control for Type I error across the three correlations, a p-value of less than .017 was required for significance.

The correlation between time spent on tasks and the number of steps for stage one was significant, \( r (134) = .78, p < .001 \); for stage two it was also significant, \( r (445) = .77, p < .001 \). This means the longer the time that a subject spent on answering a question, the greater the number of steps involved. The effect sizes were large for both stages.
The correlation between the number of steps and satisfaction for stage one was significant, \( r (134) = .51, p < .001 \); for stage two it was also significant, \( r (445) = .57, p < .001 \). This means the more steps taken to answer a question, the lower the satisfaction. The effect sizes for both stages were large.

The correlation between the time spent and satisfaction for stage one was significant, \( r (134) = .46, p < .001 \); for stage two it was also significant, \( r (445) = .61, p < .001 \). This means the more time spent on answering a question, the lower the satisfaction. The effect size for stage one was medium, for stage two it was large.

Based on those statistical results, it is concluded that there is correlation between efficiency and satisfaction. Users feel less satisfied when the task requires more steps to get to the answer.

5.12.3 Effectiveness and Efficiency

5.12.3.a Effectiveness and Steps

Two ANOVAs were conducted, for stages one and two respectively, to evaluate whether the group means of steps on correct answers and incorrect answers differ significantly from each other. The independent variable, effectiveness, included two levels: correct and incorrect answers. The dependent variable was the number of steps.

The ANOVA for stage one was significant, \( F (1, 163) = 29.19, p < .001 \). The strength of relationship between effectiveness and steps, as assessed by \( \eta^2 \), was strong, with the effectiveness factor accounting for 15% of the variance of the dependent variable.
The ANOVA for stage two was also significant, $F(1, 445) = 82.84, p < .001$. The strength of relationship between effectiveness and steps, as assessed by $\eta^2$, was strong, with the effectiveness factor accounting for 16% of the variance of the dependent variable.

As indicated in Table 39 and Figure 10, incorrect answers involved more steps while correct answers involved fewer steps. When the subjects knew how to get the answer, it took them fewer steps; while without that knowledge, they struggled.

Table 39

**Effectiveness and Steps**

<table>
<thead>
<tr>
<th>Correctness of Answer</th>
<th>Stage 1</th>
<th>Stage 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N=11$</td>
<td>$N=30$</td>
</tr>
<tr>
<td>Wrong Answer</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Correct Answer</td>
<td>8.18</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>16.97</td>
<td>1.47</td>
</tr>
</tbody>
</table>
5.12.3.b Effectiveness and Time

Two ANOVAs were conducted to evaluate whether the group means of time on correctness of answers and incorrect answers differed significantly from each other. The independent variable, the effectiveness factor, included two levels: correct answer and incorrect answer. The dependent variable was time spent on completing tasks.

The ANOVA for stage one was significant, $F(1, 163) = 15.70, p < .001$. The strength of the relationship between effectiveness and time, as assessed by $\eta^2$, was medium, with the effectiveness factor accounting for 9% of the variance of the dependent variable.

The ANOVA for stage two was also significant, $F(1, 445) = 107.44, p < .001$. The strength of the relationship between effectiveness and time, as assessed by $\eta^2$, was
strong, with the effectiveness factor accounting for 20% of the variance of the dependent variable.

Table 40 and Figure 11 indicate the mean time for correct and incorrect answers for stages one and two. When a subject is able to complete the task correctly, it takes the subject less time. When a subject does not know how to get to the correct answer, the subject often struggles.

Table 40

*Effectiveness and Time*

<table>
<thead>
<tr>
<th></th>
<th>Stage 1</th>
<th></th>
<th>Stage 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=11</td>
<td></td>
<td>N=30</td>
<td></td>
</tr>
<tr>
<td>Correctness of Answer</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Wrong Answer</td>
<td>4 min. 12 sec.</td>
<td>25 sec.</td>
<td>3 min. 53 sec.</td>
<td>2 min. 25 sec.</td>
</tr>
<tr>
<td>Correct Answer</td>
<td>2 min. 23 sec.</td>
<td>12 sec.</td>
<td>1 min. 48 sec.</td>
<td>1 min. 36 sec.</td>
</tr>
</tbody>
</table>

Figure 11. Significant Differences in Time Related to Correctness of Answers
Based on the analyses from one-way ANOVAs and correlation coefficients, we find that there exist interlocking relationships among effectiveness, efficiency, and satisfaction. Figure 2 indicates these relationships.

5.13 Learnability and Ease of Use

This section examines the relationship between learnability and ease of use. One of the measures of learnability is the quantifying element such as the number of tasks subjects can successfully perform on a new site. The data for ease of use is gathered from the Likert scale of question one in a post-test questionnaire (see Appendix D). In the 5-point Likert scale, 1 = easy to use, 5 = difficult to use.

The Pearson Product-Moment Correlation Coefficient was computed between the numbers of correct answers and the ratings of ease of use. The result of the correlational analysis show that the correlation was statistically significant, \( r (41) = -0.49, p < .001 \). This means that the more tasks subjects can complete on the new site, the higher they rate ease of use. The coefficient was large (\( r = -0.49 \)). The square of the correlation coefficient indicates that 24% of the variance of ease of use is accounted for by its linear relationship with ease of learning.

5.14 Demographic Factors on Effectiveness

The data analyses in this section combine all the data of forty-one subjects from both stages one and two to have a large enough sample size to yield accurate \( p \)-values. The unit of analysis in this section is the subject. There were a total of twenty-six
females (63%) and fifteen males (37%). Two subjects were in their teens (5%), thirteen in their 20s (32%), fourteen in their 30s (34%), six in their 40s (15%), and six in their 50s (15%). Nine were undergraduates (22%); twenty-six were master’s degree students (63%); and six were doctoral students (15%). Thirty-one majored in library and information science (76%); nine were in the field of communications (22%); and one was in neuroscience (2%). As to their ethnic background, twenty-five were White (61%); five were African-American (12%); ten were Asian or Asian-American (24%); and one was Hispanic (2%). The amount of time they had been at Rutgers University or Queens College varied: four had been at their school for less than one year (10%), twelve for one year (29%), twelve for two years (29%), seven for three years (17%), four for four years (10%). One had been there for six years (2%), and one for eight years (2%). When asked how often they used their library’s Web site, five (12%) replied that they used it once or twice a semester; six (15%) said once or twice a month, sixteen (39%) said once or twice a week, and fourteen (34%) said they used it daily. This information is displayed in Table 4.

5.14.1 Gender and Effectiveness

A one-way analysis of variance was conducted to evaluate the relationship between gender and number of correct answers (effectiveness). The independent variable, gender, included two levels: male and female. The dependent variable was the number of correct answers. The ANOVA was non-significant: $F(1, 39) = .65, p = .425$. The average of correct answers for females ($M = 11.77, SD = 1.99$) was close to those for males ($M = 11.13, SD = 3.07$).
5.14.2 Age and Effectiveness

A one-way analysis of variance was conducted to evaluate the relationship between age and effectiveness. The independent variable, age, included four levels: 10-19 (group 0, $N = 2$), 20-29 (group 1, $N = 13$), 30-39 (group 2, $N = 14$), 40-49 (group 3, $N = 6$), and 50-59 (group 4, $N = 6$). The dependent variable was the number of correct answers (effectiveness). The ANOVA was non-significant: $F(4, 36) = .67, p = .62$. The averages of correct answers for each group were: 10-19 ($M = 12, SD = 1.41$), 20-29 ($M = 12.23, SD = 1.59$), 30-39 ($M = 10.93, SD = 3.41$), 40-49 ($M = 12, SD = 1.27$), 50-59 ($M = 10.83, SD = 2.32$).

5.14.3 Status and Effectiveness

A one-way analysis of variance was conducted to evaluate the relationship between status and effectiveness. The independent variable, status, included three levels: undergraduates, master’s, and doctoral students. The dependent variable was number of correct answers. The ANOVA was non-significant: $F(2, 38) = .08, p = .925$. The averages for each group were: undergraduates ($M = 11.78, SD = 1.79$), master’s ($M = 11.42, SD = 2.63$), doctoral ($M = 11.67, SD = 2.66$).

5.14.4 Major and Effectiveness

A one-way analysis of variance was conducted to evaluate the relationship between academic major and effectiveness. The independent variable, status, included two levels: LIS major and non-LIS major. The dependent variable was number of correct answers. The ANOVA was non-significant: $F(1, 39) = .65, p = .427$. The
average for LIS students was $M = 11.71, SD = 2.53$ while for non-LIS students it was $M = 11, SD = 2.06$.

5.14.5 Ethnic Background and Effectiveness

A one-way analysis of variance was conducted to evaluate the relationship between ethnic background and effectiveness. The independent variable, ethnic background, included four levels: White ($N = 25$), African-American ($N = 5$), Asian ($N = 10$), and Hispanic ($N = 1$). The dependent variable was the number of correct answers. The ANOVA was non-significant: $F (3, 37) = 2.58, p = .068$. The averages for each group were: White ($M = 12.2, SD = 1.73$), African-American ($M = 11.8, SD = 2.17$), Asian ($M = 9.9, SD = 3.38$), and Hispanic ($M = 10$).

5.14.6 Institution Experience and Effectiveness

This section examines the relationship between the length of time a subject has been at the institution, and effectiveness. Unlike other sections using combined effectiveness scores of the Rutgers and Queens sites, this section examines the score of the home institution only. Subjects who have been at Rutgers or Queens more than four years are grouped with the four-year group. The correlation coefficient was computed between number of years and number of correct answers. The result shows that correlation was not statistically significant: $r (41) = .08, p = .634$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
5.14.7 Frequency of Using Library Web Site and Effectiveness

This section examines the relationship between the subject’s frequency of use of the library’s Web site, and effectiveness. It considers subjects’ scores at their home institution only. A one-way analysis of variance was conducted. The independent variable, frequency, included four levels: once or twice a semester ($N = 5$), once or twice a month ($N = 6$), once or twice a week ($N = 16$), and daily ($N = 14$). The dependent variable was the number of correct answers. The ANOVA was non-significant: $F(3, 37) = 1.22, p = .316$. The averages were: once or twice a semester ($M = 6.0, SD = 1.87$), once or twice a month ($M = 7.83, SD = .75$), once or twice a week ($M = 7.06, SD = 2.02$), and daily ($M = 7.29, SD = 1.27$).

5.14.8 Interaction Effects of Gender, Frequency of Use, and Years at the Institution on Effectiveness

A $2 \times 4 \times 7$ ANOVA was conducted to evaluate the effects of gender, frequency of using the library’s Web site, and the length of time at the institution, on the number of correct answers. The ANOVA indicated no significant main effects of these three factors and no significant interaction effects. The main effect for gender was $F(1, 18) = .01, p = .945$, partial $\eta^2 = .00$. The main effect for frequency of using the library’s Web site was $F(3, 18) = .12, p = .945$, partial $\eta^2 = .02$. The main effect for the length at the institution was $F(6, 18) = 3.05, p = .031$, partial $\eta^2 = .50$. The interaction between gender and years at the institution was $F(3, 18) = .57, p = .643$, partial $\eta^2 = .09$. The interaction between gender and frequency of using the library’s Web site was $F(2, 18) = .07, p = .937$, partial
\( \eta^2 = .01 \). The interaction between years at the institution and frequency of using the library’s Web site was \( F(6, 18) = .37, p = .890 \), partial \( \eta^2 = .11 \).

**5.15 Demographic Factors on Steps**

As in 5.14, the data analyses in this section examine all the data of forty-one subjects from both stages one and two in order to have a large enough sample size to yield accurate p-values.

5.15.1 Gender and Steps

A one-way analysis of variance was conducted to evaluate the relationship between gender and number of steps. The independent variable, gender, included two levels: male and female. The dependent variable was the number of steps. The ANOVA was non-significant: \( F(1, 39) = .37, p = .545 \). The average number of steps for females was \( M = 8.52, SD = 2.46 \); and for males, \( M = 8.99, SD = 2.35 \).

5.15.2 Age and Steps

A one-way analysis of variance was conducted to evaluate the relationship between age and number of steps. The independent variable, age, included four levels: 10-19 (group 0, \( N = 2 \)), 20-29 (group 1, \( N = 13 \)), 30-39 (group 2, \( N = 14 \)), 40-49 (group 3, \( N = 6 \)), and 50-59 (group 4, \( N = 6 \)). The dependent variable was the number of steps. The ANOVA was non-significant: \( F(4, 36) = 1.33, p = .277 \). The averages for number of steps for each group were: 10-19 (\( M = 7.97, SD = .90 \)), 20-29 (\( M = 8.62, SD = 2.23 \)), 30-39 (\( M = 9.72, SD = 2.91 \)), 40-49 (\( M = 7.32, SD = 1.18 \)), 50-59 (\( M = 8.05, SD = 2.16 \)).
5.15.3 **Status and Steps**

A one-way analysis of variance was conducted to evaluate the relationship between status and number of steps. The independent variable, status, included three levels: undergraduates, master’s students, and doctoral students. The dependent variable was number of steps. The ANOVA was non-significant: \( F (2, 38) = .05, p = .952 \). The averages for each group were: undergraduates \((M = 8.81, SD = 2.41)\), master’s \((M = 8.71, SD = 2.41)\), doctoral \((M = 8.41, SD = 2.77)\).

5.15.4 **Major and Steps**

A one-way analysis of variance was conducted to evaluate the relationship between academic major and number of steps. The independent variable, status, included two levels: LIS major and non-LIS major. The dependent variable was number of steps. The ANOVA was non-significant: \( F (1, 39) = .45, p = .507 \). The average for LIS students was \( M = 8.55, SD = 2.40 \), while for non-LIS students it was \( M = 9.14, SD = 2.46 \).

5.15.5 **Ethnic Background and Steps**

A one-way analysis of variance was conducted to evaluate the relationship between ethnic background and number of steps. The independent variable, ethnic background, included four levels: White \((N = 25)\), African-American \((N = 5)\), Asian \((N = 10)\), and Hispanic \((N = 1)\). The dependent variable was the number of steps. The ANOVA was non-significant: \( F (3, 37) = .53, p = .664 \). The averages for each group were: White \((M = 8.73, SD = 1.99)\), African American \((M = 8.59, SD = 3.54)\), Asian \((M = 8.94, SD = 2.90)\), and Hispanic \((M = 5.73)\).
5.15.6 Institution Experience and Steps

This section examines the relationship between the length a subject has been at the institution and the number of steps. It considers the number of steps at the subjects’ home institution site only. Subjects who have been at Rutgers or Queens more than four years are grouped with the four-year group. The correlation coefficient was computed between number of years and number of steps. The result shows that correlation was not statistically significant, \( r (41) = .10, p = .523. \)

5.15.7 Frequency of Use of Library Web Site and Steps

This section examines the relationship between the subject’s frequency of use of the home library’s Web site and the number of steps required. A one-way analysis of variance was conducted. The independent variable, frequency, included four levels: once or twice a semester \((N = 5)\), once or twice a month \((N = 6)\), once or twice a week \((N = 16)\), and daily \((N = 14)\). The dependent variable was the number of steps. The ANOVA was non-significant, \( F (3, 37) = .66, p = .583. \) The averages were: once or twice a semester \((M = 9.09, SD = 3.56)\), once or twice a month \((M = 10.48, SD = 3.95)\), once or twice a week \((M = 9.68, SD = 2.72)\), and daily \((M = 8.52, SD = 3.10)\).

5.16 Demographic Factors on Time

5.16.1 Gender and Time

A one-way analysis of variance was conducted to evaluate the relationship between gender and time used to complete tasks. The independent variable, gender, included two levels: male and female. The dependent variable was time. The ANOVA
was non-significant: $F(1, 39) = 1.46, p = .234$. The average time for females was $M = 2$ min. 19 sec., $SD = 41$ sec. while for males it was $M = 2$ min. 38 sec., $SD = 1$ min. 3 sec.

5.16.2 Age and Time

A one-way analysis of variance was conducted to evaluate the relationship between age and time used to complete tasks. The independent variable, age, included four levels: 10-19 (group 0, $N = 2$), 20-29 (group 1, $N = 13$), 30-39 (group 2, $N = 14$), 40-49 (group 3, $N = 6$), and 50-59 (group 4, $N = 6$). The dependent variable was time. The ANOVA was non-significant: $F(4, 36) = .56, p = .695$. The averages of time for each group were: 10-19 ($M = 2$ min. 27 sec., $SD = 26.87$ sec.), 20-29 ($M = 2$ min. 19 sec., $SD = 51.76$ sec.), 30-39 ($M = 2$ min. 37 sec., $SD = 57.82$ sec.), 40-49 ($M = 2$ min. 3 sec., $SD = 21.23$ sec.), 50-59 ($M = 2$ min. 35 sec., $SD = 59.45$ sec.).

5.16.3 Status and Time

A one-way analysis of variance was conducted to evaluate the relationship between status and the time used to complete tasks. The independent variable, status, included three levels: undergraduates, master's students, and doctoral students. The dependent variable was time. The ANOVA was non-significant: $F(2, 38) = .13, p = .876$. The averages for each group were: undergraduates ($M = 2$ min. 32 sec., $SD = 1$ min.), master's students ($M = 2$ min. 23 sec., $SD = 52$ sec.), doctoral students ($M = 2$ min. 30 sec., $SD = 35$ sec.).
5.16.4 Major and Time

A one-way analysis of variance was conducted to evaluate the relationship between academic major and the time used to complete tasks. The independent variable, status, included two levels: LIS major and non-LIS major. The dependent variable was time. The ANOVA was non-significant: $F(1, 39) = 1.13, p = .295$. The average for LIS students was $M = 2$ min. 21 sec., $SD = 46$ sec. while for non-LIS students it was $M = 2$ min. 41 sec., $SD = 1$ min. 4 sec.

5.16.5 Ethnic Background and Time

A one-way analysis of variance was conducted to evaluate the relationship between ethnic background and the time used to complete tasks. The independent variable, ethnic background, included four levels: White ($N = 25$), African-American ($N = 5$), Asian ($N = 10$), and Hispanic ($N = 1$). The dependent variable was time. The ANOVA was non-significant: $F(3, 37) = .70, p = .556$. The averages for each group were: White ($M = 2$ min. 19 sec., $SD = 43$ sec.), African-American ($M = 2$ min. 46 sec., $SD = 1$ min. 25 sec.), Asian ($M = 2$ min. 36 sec., $SD = 52$ sec.), and Hispanic ($M = 1$ min. 49 sec.).

5.16.6 Institution Experience and Time

This section examines the relationship between the length a subject has been at the institution and the time used to complete tasks. It examines the data of subjects’ home institution site only. Subjects who have been at Rutgers or Queens more than four years were grouped with the four-year group. The correlation coefficient was computed
between number of years and time. The result shows that correlation was not statistically significant, \( r (41) = .06, p = .716. \)

5.16.7 Frequency of Use of Library Web Site and Time

This section examines the relationship between the subject’s frequency of use of the home library’s Web site and time used to complete tasks. A one-way analysis of variance was conducted. The independent variable, frequency, included four levels: once or twice a semester \((N = 5)\), once or twice a month \((N = 6)\), once or twice a week \((N = 16)\), and daily \((N = 14)\). The dependent variable was time. The ANOVA was non-significant: \( F (3, 37) = 2.02, p = .128. \) The averages were: once or twice a semester \((M = 2\) min. 4 sec., \(SD = 23\) sec.), once or twice a month \((M = 3\) min. 2 sec., \(SD = 1\) min. 13 sec.), once or twice a week \((M = 2\) min. 33 sec., \(SD = 49\) sec.), and daily \((M = 2\) min. 11 sec., \(SD = 42\) sec.).

5.17 Gender and Satisfaction

This section combines all the data from forty-one subjects from both stages one and two to have a large enough sample size to yield accurate \(p\)-values.

5.17.1 Gender and Ease of Use

Three ANOVAs were conducted to evaluate the relationships between gender and the ratings of ease of use on the Rutgers and the Queens sites, respectively and combined. The independent variable, gender, included two levels: male and female. The dependent variable was the rating of ease of use. The ANOVA for the Rutgers site was non-
significant: \( F (1, 39) = 1.60, p = .213 \). The average of ratings of ease of use for females was \( M = 2.50, SD = 1.03 \) while for males it was \( M = 2.93, SD = 1.10 \). The ANOVA for the Queens site was also non-significant: \( F (1, 39) = .20, p = .661 \). The average of ratings of ease of use for females was \( M = 2.62, SD = .98 \) while for males it was \( M = 2.47, SD = 1.13 \). The ANOVA for Rutgers and Queens combined was not significant: \( F (1, 80) = .35, p = .554 \). The average of ratings of ease of use for females was \( M = 2.56, SD = 1.0 \) while for males it was \( M = 2.70, SD = 1.12 \).

5.17.2 Gender and Organization of Information

Three ANOVAs were conducted to evaluate the relationships between gender and the ratings of organization of information on the Rutgers and the Queens sites, respectively and combined. The independent variable, gender, included two levels: male and female. The dependent variable was the rating of organization of information. The ANOVA for the Rutgers site was non-significant: \( F (1, 38) = 1.42, p = .24 \). The average of ratings of organization of information for females was \( M = 2.08, SD = .95 \) while for males it was \( M = 2.47, SD = 1.06 \). The ANOVA for the Queens site was non-significant also: \( F (1, 38) = .14, p = .71 \). The average of ratings of ease of use for females was \( M = 2.68, SD = 1.28 \) while for males was \( M = 2.53, SD = .99 \). The ANOVA for Rutgers and Queens combined was non-significant: \( F (1, 78) = .22, p = .64 \). The average of ratings of ease of use for females was \( M = 2.38, SD = 1.16 \) while for males was \( M = 2.50, SD = 1.01 \).
5.17.3 Gender and Terminology

Three ANOVAs were conducted to evaluate the relationships between gender and the terminology ratings, on the Rutgers and the Queens sites respectively and combined. The independent variable, gender, included two levels: male and female. The dependent variable was the rating for terminology. The ANOVA for the Rutgers site was non-significant: $F(1, 38) = .06, p = .815$. The average of ratings of organization of information for female was $M = 2.12, SD = 1.05$ while for male it was $M = 2.20, SD = 1.01$. The ANOVA for the Queens site was non-significant also: $F(1, 38) = .39, p = .534$. The average of ratings of ease of use for female was $M = 2.68, SD = 1.07$ while for male it was $M = 2.47, SD = .99$. Nor was the ANOVA for Rutgers and Queens combined significant: $F(1, 78) = .08, p = .785$. The average of ratings of ease of use for females was $M = 2.40, SD = 1.09$ while for males it was $M = 2.33, SD = .99$.

5.17.4 Gender and Attractiveness

Three ANOVAs were conducted to evaluate the relationships between gender and the ratings for visual attractiveness on the Rutgers and the Queens sites respectively and combined. The independent variable, gender, included two levels: male and female. The dependent variable was the rating for visual attractiveness. The ANOVA for the Rutgers site was non-significant: $F(1, 39) = .39, p = .535$. The average of ratings of organization of information for female was $M = 2.46, SD = 1.03$ while for male was $M = 2.67, SD = .98$. Nor was the ANOVA for the Queens site significant: $F(1, 39) = .00, p = .99$. The average of ratings for ease of use for female was $M = 2.46, SD = .95$ while for male was $M = 2.47, SD = 1.13$. The ANOVA for Rutgers and Queens combined was
non-significant:  \( F (1, 80) = .21, p = .65 \). The average of ratings for ease of use for females was \( M = 2.46, SD = .98 \) while for males it was \( M = 2.57, SD = 1.04 \).

5.17.5 Gender and Mistake Recovery

Three ANOVAs were conducted to evaluate the relationships between gender and the ratings for mistake recovery on the Rutgers and the Queens sites, respectively and combined. The independent variable, gender, included two levels: male and female. The dependent variable was the rating for mistake recovery. The ANOVA for the Rutgers site was non-significant: \( F (1, 39) = .37, p = .545 \). The average of ratings for organization of information for female was \( M = 2.04, SD = 1.22 \) while for male it was \( M = 2.27, SD = 1.03 \). The ANOVA for the Queens site, too, was non-significant: \( F (1, 39) = .08, p = .78 \). The average of ratings for ease of use for female was \( M = 2.58, SD = 1.17 \) while for male was \( M = 2.47, SD = 1.25 \). The ANOVA for Rutgers and Queens combined was non-significant: \( F (1, 80) = .05, p = .83 \). The average of ratings for ease of use for female was \( M = 2.31, SD = 1.21 \) while for male it was \( M = 2.37, SD = 1.13 \).

5.18 Ethnic Background and Satisfaction

This section combines all the data from forty-one subjects from both stages one and two to have a large enough sample size to yield accurate \( p \)-values.

5.18.1 Ethnic Background and Ease of Use

Two ANOVAs were conducted to evaluate the relationships between ethnic background and the ratings for ease of use on the Rutgers and the Queens sites.
respectively. The independent variable, the ethnic background, included four levels: White ($N = 25$), African-American ($N = 5$), Asian ($N = 10$), and Hispanic ($N = 1$). The dependent variable was the rating for ease of use. The ANOVA for the Rutgers site was non-significant: $F(3, 37) = 1.22, p = .315$. The ANOVA for Queens was significant, $F(3, 37) = 3.23, p = .033$. The strength of relationship between ethnic background and the rating for ease of use for the Queens site, as assessed by $\eta^2$, was strong, with the ethnic background factor accounting for 21% of the variance of the dependable variable. Post hoc tests were not conducted because at least one group had fewer than two cases. Table 41 indicates the differences of ratings among ethnic groups on the Queens site.

Table 41

<table>
<thead>
<tr>
<th>Ethnic Background</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>2.36</td>
<td>0.860</td>
<td>25</td>
</tr>
<tr>
<td>African-American</td>
<td>2.00</td>
<td>0.707</td>
<td>5</td>
</tr>
<tr>
<td>Asian</td>
<td>3.20</td>
<td>1.229</td>
<td>10</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4.00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>2.56</td>
<td>1.206</td>
<td>41</td>
</tr>
</tbody>
</table>

Considering that the single Hispanic had a mean which was higher than the means of the other ethnic groups, an ANOVA excluding Hispanic was conducted to verify the relationship between ethnic group and the rating for ease of use on the Queens site. The overall ANOVA was significant, $F(2, 37) = 3.67, p = .035$. The strength of relationship between ethnic background and rating for ease of use for the Queens site, as assessed by $\eta^2$, was strong, with the ethnic background factor accounting for 17% of the variance of the dependable variable. Follow-up tests were conducted to evaluate pairwise differences among the means. The test of homogeneity of variance was non-significant: $p = .064$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Dunnett's C test, a multiple comparison procedure that does not require the population variance to be equal, was used. The results indicate that there were no significant pairwise differences among the means.

To have an even larger sample size to yield accurate p-values, an ANOVA on the combined ratings of the Rutgers and the Queens sites was conducted. The ANOVA was non-significant: $F(3, 78) = 2.61, p = .057$.

These statistical analyses suggest that there are probably differences in the attitude toward ease of use among various ethnic groups. A study with a larger sample size is needed to confirm these differences.

5.18.2 Ethnic Background and Organization of Information

Two ANOVAs were conducted to evaluate the relationship between ethnic background and the ratings for organization of information on the Rutgers and the Queens sites respectively. The independent variable, the ethnic background, included four levels: White ($N = 25$), African-American ($N = 5$), Asian ($N = 10$), and Hispanic ($N = 1$). The dependent variable was the rating for organization of information. The ANOVA for the Rutgers site was non-significant: $F(3, 36) = .76, p = .525$. The ANOVA for the Queens site was significant, $F(3, 36) = 4.54, p = .008$. The strength of the relationship between ethnic background and rating for organization for the Queens site, as assessed by $\eta^2$, was strong, with the ethnic background factor accounting for 27% of the variance of the dependable variable. Post hoc tests were not conducted because at
least one group had fewer than two cases. Table 42 indicates the differences on ratings among ethnic groups on the Queens site.

Table 42

*Differences among Ethnic Groups on Rating of Organization on Queens Site (1=easy to use, 5=difficult to use)*

<table>
<thead>
<tr>
<th>Ethnic Background</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>2.42</td>
<td>1.018</td>
<td>24</td>
</tr>
<tr>
<td>African-American</td>
<td>1.80</td>
<td>0.447</td>
<td>5</td>
</tr>
<tr>
<td>Asian</td>
<td>3.30</td>
<td>1.252</td>
<td>10</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5.00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>2.63</td>
<td>1.170</td>
<td>40</td>
</tr>
</tbody>
</table>

Since there was only one Hispanic subject and his mean was higher than the means for the other ethnic groups, an ANOVA excluding Hispanic was conducted to verify the relationship between ethnic group and the rating for organization of information on the Queens site. The overall ANOVA was significant, $F(2, 36) = 4.12, p = .025$. The strength of the relationship between ethnic background and the rating for organization for the Queens site, as assessed by $\eta^2$, was strong, with the ethnic background factor accounting for 19% of the variance of the dependable variable.

Follow-up tests were conducted to evaluate pairwise differences among the means. The test of homogeneity of variance was significant, $p = .036$, indicating the rejection of the null hypothesis that the error variance of the dependent variable is equal across groups. Using the Dunnett’s $C$ test, the ratings of African-American and Asian-American for the Queens site’s organization of information differed significantly from one another.

The investigator then conducted a one-way analysis of variance on the merged data of the Rutgers and the Queens sites. The ANOVA was significant, $F(3, 76) = 3.77, p = .014$. The strength of relationship between ethnic background and rating for
organization, as assessed by $\eta^2$, was strong, with the ethnic background factor accounting for 13% of the variance of the dependable variable. Post hoc tests were not conducted because at least one group had fewer than two cases. Table 43 indicates the differences in ratings among ethnic groups.

Table 43

*Differences among Ethnic Groups on Rating of Organization on Both Sites Combined (1=clear, 5=unclear)*

<table>
<thead>
<tr>
<th>Ethnic Background</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>2.23</td>
<td>0.994</td>
<td>24</td>
</tr>
<tr>
<td>African-American</td>
<td>2.10</td>
<td>0.876</td>
<td>5</td>
</tr>
<tr>
<td>Asian</td>
<td>2.90</td>
<td>1.210</td>
<td>10</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4.00</td>
<td>1.414</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>2.42</td>
<td>1.100</td>
<td>40</td>
</tr>
</tbody>
</table>

Again, an ANOVA excluding Hispanic was conducted to verify the relationship between ethnic group and the rating for organization of information on the merged data of both the Rutgers and the Queens sites. The ANOVA was significant, $F(2, 75) = 3.36$, $p = .040$. The strength of relationship between ethnic background and rating on organization, as assessed by $\eta^2$, was medium, with the ethnic background factor accounting for 8% of the variance of the dependable variable. The test for homogeneity of variance was non-significant: $p = .153$. Follow-up tests were conducted to evaluate pairwise differences among the means. Using Turkey, which assumes equal variance to control for Type I error across the multiple pairwise comparisons, the ratings of White and Asian-American for organization differed significantly from one another.
These statistical analyses of ethnic background and the ratings for organization of information suggest that different ethnic groups may have different attitudes toward organization of information.

5.18.3 Ethnic Background and Terminology

Three ANOVAs were conducted to evaluate the relationships between ethnic background and the ratings for terminology on the Rutgers and the Queens independently and together. The independent variable, the ethnic background, included four levels: White ($N = 25$), African-American ($N = 5$), Asian ($N = 10$), and Hispanic ($N = 1$). The dependent variable was the rating for terminology. The ANOVA for the Rutgers site was not significant: $F (3, 36) = 1.37, p = .269$. The ANOVA for the Queens site was also non-significant: $F (3, 36) = 2.41, p = .083$. Further, the ANOVA for Rutgers and Queens combined was non-significant: $F (3, 76) = 2.05, p = .114$.

These statistical analyses suggest that there are probably no different attitudes among different ethnic groups on terminology.

5.18.4 Ethnic Background and Attractiveness

Two ANOVAs were conducted to evaluate the relationships between ethnic background and the ratings for attractiveness on the Rutgers and the Queens sites respectively. The independent variable, the ethnic background, included four levels: White ($N = 25$), African-American ($N = 5$), Asian ($N = 10$), and Hispanic ($N = 1$). The dependent variable was the rating for attractiveness. The ANOVA for the Rutgers site
was non-significant: $F(3, 37) = 1.49, p = .233$. The ANOVA for the Queens site was significant, $F(3, 37) = 3.42, p = .027$. The strength of relationship between ethnic background and rating for attractiveness, as assessed by $\eta^2$, was strong, with the ethnic background factor accounting for 22% of the variance of the dependable variable. Post hoc tests were not conducted because at least one group had fewer than two cases. Table 44 indicates the differences on ratings among ethnic groups.

### Table 44

**Differences among Ethnic Groups on Rating of Attractiveness on Queens Site**

<table>
<thead>
<tr>
<th>Ethnic Background</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>2.16</td>
<td>0.898</td>
<td>25</td>
</tr>
<tr>
<td>African-American</td>
<td>2.40</td>
<td>1.140</td>
<td>5</td>
</tr>
<tr>
<td>Asian</td>
<td>3.10</td>
<td>0.876</td>
<td>10</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4.00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>2.46</td>
<td>1.002</td>
<td>41</td>
</tr>
</tbody>
</table>

As there was only one Hispanic, whose mean was higher than the means of other ethnic groups, an ANOVA excluding the case of Hispanic was conducted to verify the relationship between ethnic group and the rating for attractiveness on the Queens site. The ANOVA was significant, $F(2, 37) = 3.71, p = .034$. The strength of relationship between ethnic background and the rating for attractiveness, as assessed by $\eta^2$, was strong, with the ethnic background factor accounting for 17% of the variance of the dependable variable. Because the overall $F$ test was significant, follow-up tests were conducted to evaluate pairwise differences among the means. The test of homogeneity of variance was non-significant: $p = .779$. As determined by the Dunnett’s $C$ test, White and Asian-American differed significantly in their ratings for attractiveness.
A one-way analysis of variance was also conducted on the combined data of the Rutgers and Queens sites. The ANOVA was significant, \( F(3, 78) = 4.77, p = .004 \). The strength of relationship between ethnic background and the rating for attractiveness, as assessed by \( \eta^2 \), was strong, with the ethnic background factor accounting for 16% of the variance of the dependable variable. Post hoc tests were not conducted because at least one group had fewer than two cases. Table 45 indicates the differences on ratings among ethnic groups.

Table 45

<table>
<thead>
<tr>
<th>Ethnic Background</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>2.26</td>
<td>0.965</td>
<td>25</td>
</tr>
<tr>
<td>African-American</td>
<td>2.40</td>
<td>1.075</td>
<td>5</td>
</tr>
<tr>
<td>Asian</td>
<td>3.00</td>
<td>0.795</td>
<td>10</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4.00</td>
<td>1.090</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>2.50</td>
<td>0.997</td>
<td>41</td>
</tr>
</tbody>
</table>

These statistical analyses suggest that there are probably different attitudes among different ethnic groups on attractiveness.

5.18.5 Ethnic Background and Mistake Recovery

Three ANOVAs were conducted to evaluate the relationships between ethnic background and the ratings for mistake recovery on the Rutgers and Queens sites respectively and combined. The independent variable, the ethnic background, included four levels: White \((N = 25)\), African-American \((N = 5)\), Asian \((N = 10)\), and Hispanic \((N = 1)\). The dependent variable was the rating for mistake recovery. The ANOVA for the Rutgers site was non-significant: \( F(3, 37) = .77, p = .516 \). The ANOVA for the Queens
site was non-significant: $F(3, 37) = 1.79, p = .165$. The ANOVA for Rutgers and
Queens combined was non-significant as well: $F(3, 78) = .85, p = .473$.

These statistical analyses suggest that there are no differences in attitudes toward
the ratings for mistake recovery among different ethnic groups.

Table 46 is a summary of statistical significance of ethnic background on various
aspects of satisfaction. A checkmark indicates that it is statistically significant.

Table 46

*Statistical Significance between Ethnic Background and Satisfaction*

<table>
<thead>
<tr>
<th></th>
<th>Rutgers</th>
<th>Queens</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of Use</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Terminology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attractiveness</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Mistake Recovery</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.19 Summary of Findings

5.19.1 Analysis by Task

- Subjects had the lowest effectiveness rates for task 6 (find an encyclopedia
  article) and task 7 (find an e-book) on both the Rutgers and the Queens sites. This is
  consistent in both stages of experiments; it is consistent when comparing the scores of all
  subjects and the scores of Rutgers subjects searching the Rutgers site and Queens
  subjects searching the Queens site, respectively; and it is reflected in the scores for
efficiency and satisfaction (see Sections 5.2 and 5.3).
• Task 9 (find information to set up remote access) was the easiest one for all subjects. It had the highest effectiveness rate, the fewest steps required, the least amount of time used, and the highest satisfaction. This finding is consistent in both stages (see Section 5.3).

• Task 1 (locate a book) on the Rutgers site received lower effectiveness, longer time, more steps, and lower satisfaction scores than those received by the same categories on the corresponding Queens site. Thirteen participants commented that a simple title search on the Rutgers online catalog resulted in too many hits. A concise search result is more appealing (see Sections 5.3 and 5.5.1).

• The various interfaces provided by database vendors were confusing. Some subjects suggested meta-searching capability across all databases as a solution. Libraries should continue to provide helpful subject guides to assist library users in choosing the appropriate database (see Section 5.3).

• Thirteen subjects (31%) expressed various degrees of difficulty with the terminology used in both the Rutgers and the Queens sites (see Section 5.5.4).

• Library Web sites and digital library sites should be concerned with visually impaired users when it comes to the selection of colors. Red seems to be the most problematic color (see Section 5.5.5b).

• Ten subjects liked the navigation pane on the Rutgers site, which is a new feature added when Rutgers Libraries revamped its site. This navigation pane lets users conduct the most common tasks right from the home page without having to go down through layers (see Section 5.6.a.2).
• Fourteen subjects (46%) felt lost on the Rutgers site, and seventeen subjects (57%) felt lost on the Queens site. It was noted that subjects were lost more frequently on a new site than on their institution's site. The reasons for lostness include site design, navigation, tasks, lack of confidence, and difficulty with mistake recovery (see Section 5.9).

• Twenty-four subjects (71%) considered the Rutgers site easy to navigate. Nineteen subjects (56%) considered the Queens site easy to navigate. Links should be stable, self-explanatory, consistent, and easy to redeem from mistakes (see Section 5.10).

• Forty-four subjects (73%) expressed an expectation that the clicks will eventually lead them to the correct answer (see Section 5.11).

5.19.2 Learnability

• Rutgers subjects were able to begin searching the Queens library site almost immediately, and the converse was true: Queens subjects were equally facile at the Rutgers site (see Section 5.4).

5.19.3 Criteria for Ease of Use, Organization of Information, Terminology, Visual Attractiveness, and Mistake Recovery from User’s Perspective

• Subjects evaluated “ease of use” from the perspectives of “easy to get around,” “can follow directions easily,” “easy navigation,” “clear description,” “intuitive,” and “user-friendly” (see Section 5.5.2).
Subjects evaluated “organization” from the perspectives of “simple,” “straightforward,” “logical,” “easy to look up things,” and “placing common tasks upfront” (see Section 5.5.3).

Subjects evaluated “terminology” from the perspectives of “simple,” “straightforward,” “understandable,” “generic,” “label sections clearly,” “no jargon,” “clear descriptions/explanations,” and “from user’s perspective” (see Section 5.5.4).

Subjects evaluated “attractiveness” from the perspectives of “appropriate graphics,” “readability,” “appropriate color,” “not too complicated,” and “appropriate size of font” (see Section 5.5.5).

Subjects evaluated “mistake recovery” from the perspective of “easy navigation.” It was helpful to have a navigation bar and a Back button. It was also important to be able to start over. A “Help” section was suggested (see Section 5.5.6).

5.19.4 Relationships among Criteria

There are interlocking relationships among effectiveness, efficiency, and satisfaction (see Section 5.12).

Participants were less satisfied with the system when they failed to perform the tasks correctly (see Section 5.12.1).

The longer the time it took for a subject to perform a task, the more steps were involved (see Section 5.12.2).

The greater the number of steps involved in completing a task, the lower the satisfaction expressed by the subject (see Section 5.12.2).
• The longer the time spent on completing a task, the lower the satisfaction expressed by the subject (see Section 5.12.2).

• Incorrect answers took more steps/longer time, while correct answers took fewer steps/shorter time. This means that when the subject knew how to locate the answer, he/she took fewer steps and needed less time. Otherwise, the subject took more steps and spent more time trying to answer the question (see Section 5.12.3).

• There is a correlational relationship between ease of learning and ease of use (see Section 5.13).

5.19.5 Demographic Factors and Criteria

• There is no statistically significant relationship between demographic factors (gender, age, status, major, ethnic background, years at the institution, and frequency of using the library Web site) and effectiveness (see Section 5.14).

• There are no significant interaction effects between gender, years at the institution, and frequency of using the library’s Web site, on effectiveness (see Section 5.14.8).

• There are no statistically significant relationships between demographic factors (gender, age, status, major, ethnic background, years at the institution, and frequency of using the library Web site) and number of steps (see Section 5.15).

• There are no statistically significant relationships between demographic factors (gender, age, status, major, ethnic background, years at the institution, and frequency of using the library Web site) and time used to complete tasks (see Section 5.16).
• There are no statistically significant relationships between gender and the factors of satisfaction (ease of use, organization of information, terminology, visual attractiveness, and mistake recovery) (see Section 5.17).

• There are probably different attitudes among different ethnic groups on ranking satisfaction (see Section 5.18).
Chapter 6

Discussion

This chapter discusses the major findings of the study, including the interlocking relationships among effectiveness, efficiency, and satisfaction; the correlational relationship between ease of use and learnability; user lostness; click cost; and demographic factors relating to performance and satisfaction. In addition, it discusses the appropriate number of subjects for usability research of this kind.

The aim of the present study is to develop an evaluation model to assess the usability of digital libraries. Although this study selected the Web sites of two university and college libraries as test sites, the model, the methods, and the instruments designed in this study are applicable to other academic digital libraries or information systems. The specific tasks may need to be adjusted to the purpose of the particular digital library.

It is not the intention of this study to compare the performances of the subjects on these two test sites but rather to test the applicability of the model and the instruments. The scores for effectiveness, efficiency, and satisfaction, however, can be used to compare with similar sites. The combination of quantitative and qualitative approaches demonstrates some distinct advantages. Quantitative data analyses provide performance scores and enable the study of relationships such as the ones among effectiveness, efficiency, and satisfaction, as well as the relationship between ease of use and ease of learning; while qualitative data provide explanations for Likert scores, provide feedback on improvements, and help to identify what users most care about. In many cases, users are quoted directly, thus adding richness and credibility to bare statistics.
6.1 Criteria of Usability

As reviewed in Chapter 2, the dimensions of usability include useableness, usefulness, effectiveness, efficiency, satisfaction, learnability, functional correctness, memorability, mistake recovery, ease of use, ease of navigation, understandability, appropriate level of interaction, control, helpfulness, adaptability, the quality of being engaging, and flexibility.

The evaluation model proposed in this study selects four important criteria for examination: effectiveness, efficiency, satisfaction (which covers ease of use, organization of information, clear labeling and terminology, visual appearance, contents, and error corrections), and learnability. Effectiveness, efficiency, and satisfaction are covered in ISO 9241-11 (1994). This study operationalizes these three criteria and further specifies what to examine in satisfaction. Learnability is recommended by Nielsen (1993) as the most fundamental usability criterion. Examining a digital library in the light of these four criteria enables us to evaluate that library’s ability to provide accurate information quickly, easily, and satisfactorily.

The present study found that the proposed evaluation model with its suite of instruments is a well-rounded one suitable for academic digital libraries. As stated in the beginning of this chapter, the quantitative data can be used for comparison with other digital libraries while the qualitative data provide feedback for improvement.

6.2 Number of Subjects

How many subjects are needed in a study of this kind? This study recruited forty-one subjects, including twenty from Rutgers University and twenty-one from Queens
College. Each subject was asked to perform a total of fifteen tasks (nine on his/her institution’s site and six on the other site). Are these enough both to evaluate usability and to examine the relationships among effectiveness, efficiency, and satisfaction?

Nielsen and Molich (1990) found that not quite half of all major usability problems were detected with three participants. Nielsen (1993) later found that using six participants allowed the evaluator to identify the majority of important problems. Virzi (1992) found that 80% of usability problems were detected with five participants and 90% were detected with ten participants. The same findings were confirmed by Rubin (1994). Dumas and Redish (1993) recommended six to twelve participants for a typical usability test. They said that additional participants are less and less likely to reveal new information. Nielsen (2000) strongly advocated using five users in formal usability testing. He stated, “Elaborate usability tests are a waste of resources. The best results come from testing no more than five users and running as many small tests as you can afford.” These views have become industry standard. However, Spool and Schroeder (2001) found that they needed considerably more than five users to find 85% of the problems.

This study included forty-one subjects on the grounds that it is not only a usability test but also a research project to study the relationships among factors. It required more subjects to ensure sufficient statistical power. Each subject performed fifteen tasks. The unit of analysis was search. Therefore there were a total of 615 sets of scores for effectiveness, efficiency, and satisfaction. These should provide enough statistical power to analyze the relationships among effectiveness, efficiency, and satisfaction.
Another reason for using a large number of subjects is to help eliminate the investigator's judgment call as to whether a problem discovered from the usability test is user-specific or a true usability problem. A small sample size makes this difficult to determine.

6.3 Interlocking Relationships among Effectiveness, Efficiency, and Satisfaction

The present study was divided into two phases: phase one included eleven subjects and was conducted in February/March of 2004; phase two had thirty subjects and was conducted in September/October 2004, after Rutgers revamped its Web page. Statistical analyses were performed separately. The first phase had a total of 165 sets of scores. The second phase had a total of 450 sets of scores. These should provide sufficient statistical power to analyze the relationships among effectiveness, efficiency, and satisfaction. It was found in both phases that these relationships exist. Because the relationships are confirmed in both stages, the investigator does not need to perform a combined analysis using the whole 615 sets of scores.

It was found that the longer the time it took for a subject to perform a task, the greater the number of steps involved; the greater the number of steps involved in completing a task, the lower the satisfaction; the longer the time spent on completing a task, the lower the satisfaction; further, incorrect answers entailed more steps and longer time while correct answers entailed fewer steps and shorter time. This means that when the subject knew how to locate the answer, it took fewer steps and required less time. Otherwise, the subject took more steps and spent more time trying to answer the question.
The study of Frøkjær, Hertzum, and Hornbæk (2000) found that effectiveness and efficiency are either not correlated or correlated so weakly that the correlation is negligible for all practical purposes. The study of Walker et al. (1998) found user satisfaction is not determined by efficiency. The present study found that the strength of the relationship between effectiveness and steps was strong ($\eta^2 = .15$ in stage one; $\eta^2 = .16$ in stage two) and the strength of the relationship between effectiveness and time was medium ($\eta^2 = .09$) in stage one and was strong ($\eta^2 = .20$) in stage two. The strength of the relationship between effectiveness and satisfaction was strong in both stages ($\eta^2 = .26$ in stage one; $\eta^2 = .29$ in stage two). However, effectiveness, efficiency, and satisfaction are three separate criteria. Each has its specific emphasis and should be measured separately.

It is therefore concluded that the evaluation model illustrated in Figure 2 may be modified to demonstrate these inter-locking relationships among effectiveness, efficiency, and satisfaction (see Figure 12).

6.4 Ease of Learning and Ease of Use

The literature review indicated that ease of use and ease of learning are strongly related (Roberts & Moran, 1983; Whiteside, Jones, Levy, & Wixon, 1985). The present study confirms this. Subjects gave better ratings for ease of use to the new site if they could complete more tasks successfully. The coefficient was large.
Figure 12. Interlocking Relationships among Effectiveness, Efficiency, and Satisfaction
6.5 **User Lostness**

User lostness in digital libraries occurs when: users cannot identify where they are, cannot return to previously visited information, cannot go to information believed to exist, or cannot remember the key points covered (Conklin, 1987; McKnight, Dillon, & Richardson, 1991; Theng, 1997; Theng, 1999).

The present study found that 46% of participants were lost on the Rutgers site and 57% were lost on the Queens site. A closer examination found that participants felt less lost on their home institution’s site but more frequently felt lost on the new site. The qualitative data indicates that the reasons for user lostness are: (a) confusing structure of site design, (b) lack of “Back” button, (c) lack of appropriate button to start over, (d) difficulty of the particular task, and even (e) the participant’s level of confidence.

The rates of lostness on both the Rutgers and the Queens sites were much lower than the rate (73% lostness) reported by Theng, Mohd-Nasir, and Thimbleby (2000a) for the ACM Digital Library, the Networked Computer Science Technical Reference Library, and the New Zealand Digital Library.

6.6 **Navigation**

Navigation disorientation was found by Brinck, Gergle, and Wood (2002) to be among the biggest frustrations for Web users. The present study found that 71% of subjects reported that the Rutgers site was easy to navigate and 56% of subjects said the Queens site was easy to navigate. Subjects want the links to be self-explanatory and stable. Some subjects commented that the Queens site’s drop-down menu was over-sensitive and disappeared when the mouse moved. It is also important to provide an easy
route back to the home page. Consistency of navigation bars across all pages is also important. Nielsen (1998) advises that every page should allow users to search the entire site.

6.7 Click Cost

McGillis and Toms (2001) noted that “users are very reluctant to click unless they are fairly certain they will discover what they are looking for.” The present study found that 73% of the participants declared that they expect the click(s) to lead them eventually to the correct answer. As one subject remarked, “A click was supposed to imply what you could expect.” Another subject said, “I expect, but it doesn’t always.” One said, “I just hoped for the best.” Users of library Web sites come to the site to look up information. They want to be able to get to the answer easily and rapidly. Each click should get them closer to the information.

6.8 Demographic Factors and Performance

This study examines the relationships between demographic factors (including gender, age, status, academic major, ethnic background, years at the institution, and frequency of using library Web site) and effectiveness, number of steps, and time used to complete tasks. A series of ANOVAs and the Pearson Product-Moment Correlation Coefficient were conducted. It was found that the relationships are not statistically significant.

The literature review indicates that older users (55 years or older) had significantly more difficulty using the Web site than younger users (Chadwick-Dias,
McNulty, & Tullis, 2003) and that older adults are more accurate even though their task times are longer (Groff, Liao, Chaparro, & Chaparro, 1999). The present study did not confirm those findings. This may relate to the small number of subjects aged 50 or over. There were only 6 subjects (15%) over 50 years old in this study. To focus on studying age differences in usability, a larger and evenly distributed sample size (perhaps a study with half in the younger group and half in the older group) is needed.

The literature review also indicated that undergraduates took more time to complete the given task than postgraduates (Peng, Ramaiah, & Foo, 2004). In addition, the study at the University of the Pacific indicated significant differences between certain groups of majors or fields of study (Krueger, Ray, & Knight, 2004). The present research did not confirm those findings. It was found that there were no significant statistical relationships between status or academic major, and performance.

6.9 Cross-Cultural Usability

Cross-cultural usability is an interesting topic to explore, since digital libraries are a global infrastructure. The academic libraries’ Web sites selected in this study, for instance, serve a wide variety of students and faculty from a wide variety of ethnic backgrounds. This study found that ethnic background does not have a statistically significant impact on performance. However, different ethnic group seem to rate satisfaction differently, for instance, in the areas of ease of use, organization of information, and visual attractiveness. The findings were not consistent between the Rutgers and the Queens sites. It would be interesting to conduct a further study with a
larger sample size to analyze different attitudes toward satisfaction among various ethnic groups.

Barber and Badre (1998) coined the term “culturability” indicating the merging of culture and usability. The basic premise is simple: No longer can issues of culture and usability remain separate in designing for the World Wide Web. What is user-friendly for one culture can be vastly different for another culture. The same premise is applicable in digital libraries which serve users with different ethnic backgrounds and cultures.

6.10 Gender Effect on Performance and Satisfaction

The literature review indicates that there are probably gender differences in computer-related situations. For instance, Gilroy and Desai (1986) studied 270 undergraduates and 56 master’s degree students, approximately half women and half men, and found that female college students had significantly higher computer anxiety than male students. Other research also found that women experience higher anxiety than men in using computers in general (Frankel, 1990; Lowe & Krahn, 1989). Shashaani (1994), in a study of 902 boys and 828 girls in a secondary school, found that male computer experiences were stronger when compared with females, and correspondingly boys showed more positive attitudes towards computers than girls did. However, Gefen and Straub (1997) found that, although women and men differ in their perceptions of e-mail as a mode of communication, there was no difference in either gender’s ability to use e-mail. The usability testing at Nanyang Technological University in Singapore (Peng, Ramaiah, & Foo, 2004) also indicates that there were no significant gender differences in
users’ ability to use their gateway to electronic media services or their OPAC. There have been studies reporting that the gender gap is beginning to narrow (Livingston, Maxfield, Atterberry, & Portis, 1990; Pope-Davis & Twing, 1991; Rocheleau, 1995).

The present study found that gender has statistically significant differences neither in performance, nor in the ratings for satisfaction.

A further examination was conducted of the interaction of gender, years at the institution, and frequency of using the library’s Web site, with effectiveness, to determine if a gender gap was mediated by years at the institution and frequency of using the library’s Web site. It found no significant interaction effects.

6.11 Terminology

It was found that thirteen subjects (31%) expressed various degrees of difficulty with the terminology used on both the Rutgers and the Queens sites. The concerns were centered on the assumption of the library sites that users have a common sense of library terms, and the need users feel for better description/explanation of library jargons. Perhaps libraries could consider the addition of a glossary section or explanatory phrases to their online catalogs to explain the meaning of all common library terms. This practice has been carried out at the National Electronic Library for Communicable Diseases (Williams, Madle, Weinberg, Kostkova, & Mani-Saada, 2004) where medical terms are inevitable. California State University’s San Marcos Library also reports that it uses “explanatory roll-overs” to explain library jargons (Thompson, 2003). The provision of a glossary would go some way toward simplifying and explaining the language used, to
ensure that precise detail is not lost. The use of explanatory roll-overs or pop-up windows to provide definitions would also be helpful.

Twelve subjects (29%) complimented the terminology on the Rutgers and Queens sites. The comments from participants reflected the kinds of terminology desired by users. For example, the terms need to be simple, straightforward, generic, free of library jargon, and easy for users to understand.

The precise use of words and a common terminology will lead to higher usability of the system.

6.12 Performance Data

The literature review has indicated that there is a need for benchmarks for comparisons between effectiveness and efficiency. This is not to determine whether one system is better than another, but rather to serve as an indicator of whether or not a function of a system is well-implemented. Tables 47, 48, and 49 provide a glance at the results of effectiveness and efficiency tests at the University of the Pacific, the University of Illinois at Chicago, Rutgers University, and Queens College. The results indicate that the functions of locating books, locating journals, locating instructions for setting up remote access, and using databases at both the Rutgers and the Queens sites are well implemented. There is a need for more statistics to compare other functions. Table 47 shows the percentage of correctly completed tasks.
Table 47

*Results of Effectiveness Test*

<table>
<thead>
<tr>
<th>Function</th>
<th>U. of the Pacific</th>
<th>U. of Illinois at Chicago</th>
<th>Rutgers (Stage 1)</th>
<th>Rutgers (Stage 2)</th>
<th>Queens College (Stage 1)</th>
<th>Queens College (Stage 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locate a book</td>
<td>91%</td>
<td>80%</td>
<td>100%</td>
<td>83%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Locate a journal</td>
<td>50%</td>
<td>58%</td>
<td>100%</td>
<td>80%</td>
<td>83%</td>
<td>93%</td>
</tr>
<tr>
<td>Locate instructions on remote access</td>
<td>58%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Use of database</td>
<td>75%</td>
<td>93%</td>
<td>62%</td>
<td>83%</td>
<td>84%</td>
<td></td>
</tr>
</tbody>
</table>

Table 48 compares the average times used in locating a journal at the University of Illinois, Rutgers, and Queens College. The other tasks either are not comparable or are not reported in the literature and therefore are not included in the table. Table 49 compares the average number of steps required at these three institutions. These two tables indicate that the function of locating journals at both the Rutgers and the Queens sites is well-implemented.

Table 48

*Results of Efficiency Test in Terms of Time*

<table>
<thead>
<tr>
<th>Function</th>
<th>U. of Illinois at Chicago</th>
<th>Rutgers (Stage 1)</th>
<th>Rutgers (Stage 2)</th>
<th>Queens (Stage 1)</th>
<th>Queens (Stage 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locate a journal</td>
<td>1 min. 48 sec.</td>
<td>1 min. 9 sec.</td>
<td>1 min. 39 sec.</td>
<td>1 min. 5 sec.</td>
<td>1 min. 26 sec.</td>
</tr>
</tbody>
</table>
Table 49

Results of Efficiency Test in Terms of Steps

<table>
<thead>
<tr>
<th>Function</th>
<th>U. of Illinois at Chicago</th>
<th>Rutgers (Stage 1)</th>
<th>Rutgers (Stage 2)</th>
<th>Queens (Stage 1)</th>
<th>Queens (Stage 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locate a journal</td>
<td>11</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

6.13 Other Observations

It was observed that some users prefer to use the library Web site’s internal search engine rather than trying to figure out how to navigate the site. Both the Rutgers and the Queens sites provided a site search engine to locate information. Several subjects (R2-1, R2-3, Q2-6, and Q2-9) commented that the best feature was the site search.

Several participants said that they would give up trying to locate the information on their own; they would rather call a librarian for help.

It was found that participants had more difficulty finding an encyclopedia article and finding an e-book on both sites. This difficulty was consistent in both stages one and two, and is reflected in the effectiveness, efficiency, and satisfaction scores.

6.14 Other Discussion

Participants in this study provided many constructive comments on ways to improve the usability of the Rutgers and the Queens sites, including comments on graphic design, color, the placements of buttons, navigation, the use of explanatory texts, and other things. It appears that an attractive Web site motivates users to perform better.
Attractive sites help to create positive emotions that lead to better thinking and performance.

It should be noted that library Web sites are constantly being changed. Some of the improvements suggested by participants in this dissertation study have since been made on the two libraries’ Web sites. It should also be noted that some comments are a reflection of the participants’ ignorance. For example, one Queens participant commented that the Rutgers site does not provide a browse function on its online catalog when in fact the Rutgers online catalog allows you to choose a “begins with” option, which is essentially the same as the Queens site’s “browse” feature.

When an information system fails, there may be technical problems or there may be a user’s behavioral problem, or there may simply be user education issues. This study evaluated information systems using real tasks, real users, and a real environment, in order to validate the systems against the user’s actual task needs, which were reflected in the nine representative tasks.

This study treated user satisfaction apart from user performance (that is, effectiveness and efficiency) but found that they correlate.

In addition to the three ISO-defined usability factors (i.e., effectiveness, efficiency, and satisfaction), this study examined learnability, which is viewed by Nielsen (1993) as the most fundamental usability attribute. The present study provides operational criteria for the measurement of learnability, effectiveness, efficiency, and satisfaction.

The evaluation model proposed in this study examines performance measurements in terms of accuracy rates, time, and steps; it uses Likert scales ratings for satisfaction; and asks participants to express opinions regarding overall reaction, best
features, worst features, desired features, and suggestions/comments. These instruments
together evaluate usability wholly, both inherent and apparent, covering outcome, process,
and task attributes.
Chapter 7

Conclusions

This chapter presents conclusions, contributions, limitations, as well as suggestions for future research.

7.1 Conclusions

The goals of this dissertation study were: to develop a model of evaluating usability for academic digital libraries; to use the model for the development of measurements, instruments, and methods for evaluating academic digital libraries; to test the measurements, instruments, and methods by applying them in two academic digital libraries; to generalize the model, instruments, and methods for use in academic libraries; and to study the relationships among effectiveness, efficiency, and satisfaction. These objectives have been achieved. A model and a suite of instruments have been designed, tested, and modified. Criteria have been established for the evaluation of digital libraries. The model proposed in this study can be applied to other academic digital libraries or information systems, although specific tasks need to be tailored to each specific system. The study found interlocking relationships among effectiveness, efficiency, and satisfaction (see Figure 12). As might be expected, it also found a correlational relationship between ease of use and learnability. In addition, it explored the user lostness and click costs issues and established the users’ criteria for usability factors: ease of use, organization of information, terminology, attractiveness, and mistake recovery. The users’ comments are especially helpful in establishing those features most important to digital library users.
The digital library is an information system over a network that is organized and well-managed, and that supports the creation, use, and searching of digital objects. It is important that we recognize the digital library as a tool that supports the user's information search. Users are looking for an information system that is easy and intuitive to use. This study applied user-centered formal usability testing techniques to measure usability from the perspectives of effectiveness, efficiency, satisfaction, and learnability, while considering both performance elements and satisfaction.

As discussed in Chapter 2, usability is a multidimensional construct that covers inherent usability and apparent usability; involves outcome, process, and task; and is technical, cognitive, social, and design-oriented. It is impractical to examine all aspects of usability in one study. The model presented in this dissertation is a well-rounded one. This study chooses to operationalize the three criteria discussed in ISO definitions (effectiveness, efficiency, and satisfaction), as well as learnability. The evaluation model proposed in this dissertation considers both the quantifying elements (time, accuracy rate, number of steps to complete tasks) and the subjective criteria of satisfaction, which is further examined in the areas of ease of use, organization of information, labeling, visual appearance, content, and error correction. Although the model was tested in two academic libraries in this study, it is also applicable to other types of digital libraries or information systems.

Effectiveness in this model is measured by accuracy rate. Efficiency is measured by the speed and the number of steps needed to complete tasks. Satisfaction, as described earlier, is multi-faceted and is examined in the areas of ease of use, organization of information, labeling, visual appearance, content, and error correction. The use of Likert
scales and a questionnaire for comments provide both quantitative and qualitative data. Learnability is examined from three aspects: the time required before the subject can begin to perform a task at a new site, the accuracy rate of completed tasks, and the speed of completing tasks. The study found that the proposed model works. The instruments were revised slightly after stage one to add studies of user-lostness, navigation, and click cost.

The evaluation technique used in this study was primarily formal usability testing, supplemented by a post-test questionnaire, interview, observation, and think-aloud protocol. It used real users to conduct real tasks.

The study employed data triangulation across time, space, and persons to establish factual accuracy. The study had a total of forty-one subjects; it was divided into two stages, and the experiments were conducted on more than one site. This was to verify the interlocking relationships among usability factors -- effectiveness, efficiency, and satisfaction; to measure learnability by asking participants to use a new site; and to test cross-institutional usability.

The following are the major findings of this study:

**Performance Data**

One outcome of the study has been the provision of scores from the effectiveness and efficiency tests. These numbers can be used for comparison with the scores for similar sites to provide an indication of their levels of usability.
Interlocking Relationships among Effectiveness, Efficiency, and Satisfaction

The present study found interlocking relationships among effectiveness, efficiency, and satisfaction. The longer the time required performing a task, the greater the number of steps involved. The greater the number of steps involved, the lower the satisfaction. The longer the time spent on completing a task, the lower the satisfaction.

The strength of the relationship between effectiveness and the number of steps was strong. The strength of the relationship between effectiveness and time was medium in stage one and was strong in stage two. The strength of the relationship between effectiveness and satisfaction was strong in both stages. However, the study recognizes that effectiveness, efficiency, and satisfaction are three separate criteria. Each has its specific emphasis and should be measured separately.

Relationship between Ease of Use and Ease of Learning

This study found that ease of use and ease of learning are strongly correlated. A site that is easy to learn receives a better rating for ease of use.

Effect of Demographic Factors on Performance and Satisfaction

The study results indicate that demographic factors, including gender, age, status, academic major, ethnic background, years at the institution, and frequency of using the library Web site, do not have any statistically significant effect on performance. However, it appears that ethnic background may affect satisfaction ratings.
**User Lostness**

The study found that 46% of its participants got lost on the Rutgers site and 57% got lost on the Queens site. A closer examination found that participants felt less lost on their home institution’s site but more frequently felt lost on the new site. The qualitative data indicates that the reasons for user lostness are: (a) confusing structure of site design, (b) lack of “Back” button, (c) lack of appropriate button to start over from home page, (d) difficulty of the particular task, and even (e) participant’s level of confidence.

**Click Cost**

Seventy-three percent of the participants in this study declared that they expect the click(s) to lead them eventually to the correct answer. Users of library Web sites come to a site to look up information. They want to be able to get to the answer easily and rapidly. Each click should get them closer to the information.

**Terminology**

The study found that 31% of the participants expressed concerns about the terminology used on the Rutgers and the Queens libraries’ Web sites. The library sites and the users did not share a common understanding of library terms, and the users felt the need of better descriptions/explanations of the technical vocabulary of the library.
User's Perspectives on Ease of Use, Organization, Terminology, Visual Attractiveness, and Mistake Recovery

A usability test is a process of learning from users. This dissertation study solicited the users' perspectives on ease of use, organization, terminology, visual attractiveness, and mistake recovery. The subjects evaluated ease of use from the perspectives of “easy to get around,” “can follow directions easily,” “easy navigation,” “clear description,” “intuitive,” and “user-friendly.” The subjects evaluated organization from the perspectives of “simple,” “straightforward,” “logical,” “easy to look up things,” and “placing common tasks up front.” They evaluated terminology from the perspectives of “simple,” “straightforward,” “understandable,” “generic,” “label sections clearly,” “no jargon,” “clear descriptions/explanations,” and “from user’s perspective.” They evaluated attractiveness from the perspectives of “appropriate graphics,” “readability,” “appropriate color,” “not too complicated,” and “appropriate size of font.” The subjects evaluated mistake recovery from the perspectives of “easy navigation,” “easy to start over,” and “online help.”

Implications of This Research

The findings of this study have implications for evaluating a specific digital library and for designing digital libraries generally. Digital library development should put the users first. This study is an empirical, exploratory, user-centered evaluation that considers effectiveness, efficiency, satisfaction, and learnability – a matrix of four important criteria in usability assessment. The satisfaction criterion is further examined in the areas of ease of use, organization of information, labeling and terminology, visual

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
appearance, content, and error correction. The quantifying elements, the comments received from participants, and the experimenter's observation provide valuable data to enhance the usability of the test sites and to serve as a good reference for other digital libraries or information systems.

7.2 Limitations

The model proposed in this dissertation study is applicable to other information systems. The instruments should be modified to fit the specific application.

One limitation of this research is the lack of a specific task to evaluate the digital projects available on the Rutgers Libraries site. Those projects are developed as part of A Bridge to the Future: The Rutgers Digital Library Initiative (http://www.libraries.rutgers.edu/rul/about/long_range_plan.shtml). In the original 2002 pilot study for the present project, one task involved searching one of the digital projects/collections available through the Rutgers site. This task was eliminated in the formal experiments for this study in order to keep the tasks generic and applicable to other academic digital libraries such as the Queens site.

Another limitation of this study is the sample size. Although forty-one subjects is a large enough sample for formal usability testing and to examine relationships among various factors, it may not be large enough to examine cross-cultural usability or age-specific differences. We would need more participants in each ethnic group to investigate cultural influences on preferences. We would need a larger sample and more balanced age distribution to study the relationship between age and usability.
7.3 Future Research

Future research can extend from this study to examine learnability further. Learnability can be defined as ease of learning and retention of learning over time. This study focuses on ease of learning and examines the quantifying elements of learnability: 1) how soon the participants are able to begin to use the new site, and 2) at what level they are able to achieve proficiency. The future study could be designed so as to recall subjects after a period of time to test their memory of certain functions. Future research could also be designed so as to measure specifically the learning effort required to master certain functions of an information system.

Cross-cultural usability is another area worthy of future research. Digital libraries available via Internet serve a wide array of diverse users. The present study found that ethnic background may have some influence on the ratings subjects give to satisfaction, particularly in the areas of ease of use, organization of information, and visual attractiveness. More subjects in each ethnic group are needed for future research on this topic.

7.4 Contributions

The contributions of this dissertation study are: (a) the provision of an evaluation model for digital libraries; (b) the provision of performance data from two academic sites; (c) the demonstration of interlocking relationships among effectiveness, efficiency, and satisfaction; (d) the finding of a correlational relationship between learnability and ease of use; (e) the establishment of operational criteria and strategies to measure effectiveness, efficiency, satisfaction, and learnability; (f) the identification of the causes of user
lostness; (g) the identification of factors that contribute to ease of navigation; (h) the confirmation of click cost; (i) the establishment of the fact that the demographic factors of gender, age, status, academic major, ethnic background, years at the institution, and frequency of using the library’s Web site do not have a statistical significance on performance; (j) the indication that ethnic background may affect satisfaction ratings; (k) the identification of users’ criteria for evaluating ease of use, organization, terminology, attractiveness, and mistake recovery; (l) a review of the way usability has been and should be defined in the context of the digital library; and (m) a review of the usability evaluation methods that have been applied in academic digital libraries.
Appendix A-1.
Participant Information and Consent Form for Rutgers Subjects

Purpose
You are invited to participate in this experiment to study usability of two library Web sites. The purpose of this study is to evaluate the products and not your capabilities.

Task Description
1. You will be asked to complete a pre-search questionnaire.
2. You will be asked to perform a series of tasks on two Web sites.
3. You will be asked to “think aloud” about what you are doing and why you are doing it during your session. What you say will be recorded for later analysis.
4. Your screen activity such as search history and mouse/keyboard movement will be captured so that we may review the activities later. You should note that the videotape that we make will be only of the monitor screen and your voice, but it is possible that the back of your head might at times appear on the tape.
5. The duration of your participation will be about one hour and a half.
6. Do not worry if you can’t find correct answers. This is not a test on you.
7. At the completion of the tasks, you will be asked to answer a post-search questionnaire.

Waiver
By signing this form, you give your consent to Judy Jeng, the Ph.D. candidate, to use your voice, verbal statements, and videotape pictures, but not your name, for evaluation of the Web site.

Voluntary Participation
Participation in this study is voluntary. You may leave at any time although it will be much appreciated if you can stay to complete all tasks.

Confidentiality
All data generated in this study, including videotapes and questionnaires, will be kept confidential. Any public presentations or published reports of the results of this study will not include any means of identifying you.

Risks
To the best of our knowledge, there are no physical or psychological risks associated with participating in this study.

If you agree with these terms, please indicate your acceptance by signing below.

Signed ____________________________ Date __________________

Your name (please print): ____________________________
Appendix A-2
Consent Form for Queens Subjects

Queens College/CUNY
65-30 Kissena Blvd.
Flushing, NY 11367

CONSENT TO SERVE AS A PARTICIPANT IN A RESEARCH PROJECT

Project Title: Usability of the Digital Library: An Evaluation Model

Project Director/Investigator: Colleen Cool, Associate Professor, Graduate School of Library and Information Studies, Queens College, 718-997-3788

Research/Study Investigator: Judy Jeng, Ph. D. Candidate, School of Communication, Information, and Library Studies, Rutgers University, 201-200-2372

You are being asked to participate in a research project conducted through Queens College CUNY and Rutgers University. If you decide to participate, Queens College requires that you give your signed authorization to participate in this research project.

A basic explanation of the project is written below. Please read this explanation and discuss it with the Research Investigator. If you then decide to participate in the research project, please sign the last page of this form.

Nature and Purpose of the Project:
The proposed research is to develop and evaluate methods and instruments for assessing usability of digital libraries. A model of usability evaluation is proposed and a suite of instruments is designed. These are to be tested on Rutgers Libraries Web site and Queens College Web site. The purpose of this study is to evaluate the products and not your capabilities.

Explanation of Procedures:
1. You will be asked to complete a pre-search questionnaire.
2. You will be asked to perform a series of tasks on the two Web sites.
3. You will be asked to “think aloud” about what you are doing and why you are doing it during your session. What you say will be recorded for later analysis.
4. Your screen activity such as search history and mouse/keyboard movement will be captured so that we may review the activities later. You should note that the videotape that we make will be only of the monitor screen and your voice, but it is possible that the back of your head might at times appear on the tape.
5. The duration of your participation will be about one hour and a half.
6. Do not worry if you can’t find correct answers. This is not a test on you.
7. At the completion of the tasks, you will be asked to answer a post-search questionnaire.

Potential Discomfort and Risks:
To the best of our knowledge, there are no physical or psychological risks associated with participating in this study.

Potential Benefits:
You will have opportunity to use Rutgers Libraries Web site and learn how other institution organize their online resources.

Costs/Reimbursements:
You will be compensated $10 for participating in this study.

Alternatives to Participation:
N/A

Termination of Participation:
Your cooperation in completing the tasks and think aloud is expected. If you can’t think aloud, the investigator may terminate your session. Do not worry if you can’t find correct answers. This is not a test on you.

Confidentiality:
All data generated in this study, including videotapes and questionnaires, will be kept confidential. Any public presentations or published reports of the results of this study will not include any means of identifying you.

Withdrawal from the Project:
Your participation in this research project is completely voluntary. You may decide to stop participating in this project at any time. However, if you leave before completing your searches, the investigator will not compensate you.

Who to call if you have any questions:
The approved stamp on this consent form indicates that this project has been reviewed and approved for the period indicated by the Queens College (CUNY) Institutional Review Board for the Protection of Human Subjects in Research and Research Related Activities.

If you have any questions about your rights as a research participant, or to report a research related injury, you may call:

Office of Research and Sponsored Programs, Queens College (CUNY),
Telephone # (718) 997-5400.
If you have concerns or questions about the conduct of this research project you may call:

Colleen Cool  
Associate Professor  
Graduate School of Library and Information Science  
Queens College (CUNY)  
Flushing, NY 11367  
Telephone: 718-997-3788

What signing this form means:

By signing this consent form, you agree to participate in this research project. The purpose, procedures to be used, as well as the potential risks and benefits of your participation have been explained to you in detail. You can refuse to participate or withdraw from this research project at any time without penalty. Refusal to participate in this study or withdrawal from this study will have no effect on any services you may otherwise be entitled to from Queens College (CUNY). You will be given a copy of this consent form.

Printed Name of Participant

______________________________    __________________________
Participant Signature               Today’s Date

Printed Name of Research/Study Investigator

______________________________    __________________________
Signature of Research/Study Investigator               Today’s Date
Appendix B. Pre-Test Questionnaire

Thanks very much for agreeing to participate in this experiment. All of your personal data that we collect will be entirely confidential, viewed only by the experimenter, and shared only as part of group results. But first, we would like to gather a bit of background information about you, so that we will be better able to interpret your use of and reactions to the system.

Participant #

Date: _________________

Gender: ___ Male ___ Female

Age: _____

What is your current status:

_____ Undergraduate       _____ Master’s Student

_____ Doctoral Student

Major/Department: _____________________________________

How many years have you been at Rutgers or Queens? _______

If you are from foreign country, how long have you been in the U.S.? ________ years

Your original nationality: _________________

Ethnic group: ___ White ___ African American ___ Asian

___ Hispanic ___ Native American ___ Other: ____________

How often do you use the Library’s Web site:

_____ Never used it

_____ Once or twice a semester

_____ Once or twice a month

_____ Once or twice a week

_____ Daily
Appendix C. Usability Testing Questions

The goal of this test is to evaluate the usability of the library’s Web site. I will ask you a series of questions and would like you to think out loud while you look for the answer. Some questions are easy and some are more difficult. Do not worry if you can’t find the answer every time. Please remember that we are testing the effectiveness of the site design and this is not a test of you. The whole test should take less than an hour. I thank you.

1. Does the library have a copy of *Gone with the Wind*, book format, by Margaret Mitchell?

   Please rank from 1 to 5 regarding the ease of use of the system, 1 being the easiest and 5 being the most difficult.

   1 2 3 4 5
   Easy to use Difficult to use

   Your comment: ____________________________________________________________

   __________________________________________________________

2. Does the library currently subscribe to paper copy of *Advertising Age*?

   1 2 3 4 5
   Easy to use Difficult to use

   Your comment: ____________________________________________________________

   __________________________________________________________

3. Use a database to find an article about nursing homes and mental illness.

   1 2 3 4 5
   Easy to use Difficult to use

   Your comment: ____________________________________________________________

   __________________________________________________________
4. **Use a database to find a journal article on gospel music.**

   1  2  3  4  5
   Easy to use  Difficult to use

   Your comment: 

   

5. **I am interested in investing in what are referred to as “callable securities.” Please use a database to find a recent article about them.**

   1  2  3  4  5
   Easy to use  Difficult to use

   Your comment: 

   

6. **Find an encyclopedia article about French wine.**

   1  2  3  4  5
   Easy to use  Difficult to use

   Your comment: 

   

7. **Find an e-book called “The story of mankind.”**

   1  2  3  4  5
   Easy to use  Difficult to use

   Your comment: 

   

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
8. Can alumni enjoy inter-library loan service?

   1  2  3  4  5
   Easy                     Difficult

   Your comment: ________________________________

9. Find instruction on how to set up your home computer to have remote access to the library electronic resources.

   1  2  3  4  5
   Easy to find                Difficult to find

   Your comment: ________________________________
Appendix D-1. Post-Test Questionnaire (1st Stage)

Thanks again for participating in this experiment. This questionnaire gives you an opportunity to tell us your reactions to the system you used. Please circle a number on the scale to indicate your reactions. Please write comments to elaborate on your answers. I will go over your answers with you to make sure that I understand all of your responses. Thank you.

1. Please rate the ease of use of the Web site.

1  2  3  4  5
Easy  Difficult

Your comment: ________________________________________

2. What do you think about the organization of information on the site?

1  2  3  4  5
Clear  Unclear

Your comment: ________________________________________

3. What do you think about the terminology used in the site? Are categories clearly labeled?

1  2  3  4  5
Clear  Unclear

Your comment: ________________________________________

4. Is the site visually attractive?

1  2  3  4  5
Attractive  Unattractive

Your comment: ________________________________________
5. What is the best feature(s) of the site?

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

6. What is the worst feature(s) of the site?

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

7. What new content or features that you would like to see on the site?

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

8. Can you recover from mistakes easily?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Easy</td>
</tr>
</tbody>
</table>

Your comment: ________________________________________________________________

________________________________________________________________________
9. Your overall reaction to the system:

   1  2  3  4  5
   Satisfied  Unatisfied

Your comment: __________________________________________________________

_____________________________________________________________________

10. Do you have any other comments about the Web site?

_____________________________________________________________________

_____________________________________________________________________
Appendix D-2. Post-Test Questionnaire (2nd Stage)

Thanks again for participating in this experiment. This questionnaire gives you an opportunity to tell us your reactions to the system you used. Please circle a number on the scale to indicate your reactions. Please write comments to elaborate on your answers. I will go over your answers with you to make sure that I understand all of your responses. Thank you.

1. Please rate the ease of use of the Web site.

   1  2  3  4  5
   Easy Difficult

   Your comment: ____________________________________________

2. What do you think about the organization of information on the site?

   1  2  3  4  5
   Clear Unclear

   Your comment: ____________________________________________

3. What do you think about the terminology used in the site? Are categories clearly labeled?

   1  2  3  4  5
   Clear Unclear

   Your comment: ____________________________________________

4. Is the site visually attractive?

   1  2  3  4  5
   Attractive Unattractive

   Your comment: ____________________________________________
5. What is the best feature(s) of the site?

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

6. What is the worst feature(s) of the site?

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

7. What new content or features that you would like to see on the site?

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

8. Can you recover from mistakes easily?

   1  2  3  4  5
   Easy       Difficult

   Your comment: __________________________________________________________
________________________________________________________________________
9. Your overall reaction to the system:

1  2  3  4  5
Satisfied  Unsatisfied

Your comment: ____________________________________________

10. Do you feel lost while using the site?

_____ Yes  _____ No

Your comment: ____________________________________________

11. Is the site easy to navigate?

_____ Yes  _____ No

Your comment: ____________________________________________

12. When you click a button on the Web page, do you expect that the click will lead you to correct answer?

_____ Yes  _____ No

Your comment: ____________________________________________

13. Do you have any other comments about the Web site?

_________________________________________________________
References


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CURRICULUM VITA

Judy H. Jeng

EDUCATION

Ph. D. 2006 Rutgers University
School of Communication, Information and Library Studies

MLIS 1984 University of Texas at Austin
Graduate School of Library and Information Science

BA 1981 National Taiwan University
School of Library Science

1982 Certificate in Computer Programming
National Taiwan University, Graduate School of Electrical Engineering

PUBLICATIONS


“What is usability in the context of the digital library and how can it be measured?” Information Technology and Libraries, 24(2), June 2005, pp. 47-56.


"Artificial Intelligence and Authority Control: Together at Last?" LITA Newsletter 15, no.4 (Fall 1994): 28-29.


**Book and Software Reviews**


EXPERIENCE
Head of Collection Services, New Jersey City University. (2005-present).
Acquisitions Librarian. National Taiwan University (1982).
Research Assistant. National Taiwan University. CCCII Project (1981).