Understanding Usefulness in Human-Computer Interaction to Enhance User Experience Evaluation

A Thesis

Submitted to the Faculty of Drexel University by Craig Matthew MacDonald in partial fulfillment of the requirements for the degree of Doctor of Philosophy April 2012
Dedications

For Lily, who believed in me even when I doubted myself.
Acknowledgements

Even though writing a dissertation can be a somewhat solitary and lonely process, it cannot be done without the kindness, support, and generosity of many people. I'd like to thank those people now.

First and foremost, I'd like to thank my advisor, Michael Atwood, for challenging me when I needed to be challenged and pushing me when I needed to be pushed. Your perspective and approach to HCI will be forever ingrained in my own research, and for that I thank you. You’ve definitely earned your retirement. I'd also like to thank the members of my committee: Susan Wiedenbeck, Michelle Rogers, Denise Agosto, and Kasper Hornbæk. You helped me think like a researcher, and it was truly an honor working with all of you.

Thank you to the faculty and staff at the iSchool at Drexel, especially Eileen Abels for being a great mentor (and turning me into a library person). Special thanks also to Marie Fazio and Brenna Pellegrini – what we do as Ph.D. students would literally never get done without the two of you – and all of the doctoral students who volunteered their time and energy to provide me with feedback and support.

Thank you to my family. You may not have had any idea what I was doing, but you offered me encouragement and support anyway. Thanks especially to my parents, Stan and Melanie MacDonald, for never making me stop going to school. This is the last graduation, I promise. Lily, you’ve given me more support and encouragement than I probably deserve and for that I’ll always be grateful. Thanks also to my brothers, brothers- and sisters-in-law and my nieces and nephew, especially Megan (age 5) who told me to “get what I get and don't get upset” and Grace (age 6) who told me that “you can do it, never give up.” You probably weren’t talking about finishing a dissertation, but I took your advice anyway.
TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................... ix
LIST OF FIGURES ................................................................................................................ x
ABSTRACT ............................................................................................................................ xii

1. INTRODUCTION .............................................................................................................. 1

2. LITERATURE REVIEW ...................................................................................................... 6

2.1 History of Evaluation in HCI .......................................................................................... 6

2.1.1 The Reliability Phase (1950s and before) ..................................................................... 9

2.1.2 The System Performance Phase (1950s-1960s) ............................................................. 9

2.1.3 The User Performance Phase (1960s-1970s) ............................................................... 10

2.1.4 The Usability Phase (1980s-2000s) .............................................................................. 14

2.1.5 The User Experience (UX) Phase (2000s-Present) ......................................................... 17

2.1.6 Summary ..................................................................................................................... 20

2.2 Defining Usefulness ........................................................................................................ 21

2.2.1 Functions ..................................................................................................................... 23

2.2.2 Tasks ............................................................................................................................ 23

2.2.3 User Goals .................................................................................................................... 25

2.2.4 Use Context .................................................................................................................. 25

2.2.5 Summary ..................................................................................................................... 27

2.3 Usefulness and Context .................................................................................................. 27

2.3.1 Studies of Usefulness and Context .............................................................................. 28

2.3.2 Defining and Understanding Context ......................................................................... 31

2.3.3 The Elements of Context ............................................................................................ 34
2.3.4 Summary ...................................................................................................................47
2.4 Usefulness and Usability..............................................................................................47
2.4.1 Defining and Measuring Usability........................................................................7
2.4.2 Usability, Perceived Ease of Use and Perceived Usefulness.................................50
2.4.3 The Relationship between Usability and Usefulness.............................................53
2.4.4 Summary ...................................................................................................................57
2.5 Usefulness, Aesthetics and Enjoyment .....................................................................58
2.5.1 Aesthetics ..................................................................................................................58
2.5.2 Enjoyment .................................................................................................................60
2.5.3 Summary ...................................................................................................................62
2.6 Summary and Research Questions...........................................................................62
2.7 Significance ................................................................................................................64
3. RESEARCH METHODS ..............................................................................................66
3.1 Introduction ................................................................................................................66
3.2 Variables ....................................................................................................................67
3.2.1 Independent Variables ............................................................................................67
3.2.2 Dependent Variables ...............................................................................................72
3.2.3 Summary ...................................................................................................................76
3.3 Experimental Design ..................................................................................................77
3.4 Experimental Procedure ............................................................................................79
3.5 Manipulation Check ...................................................................................................81
3.6 Data Collection Instruments .......................................................................................81
3.6.1 Post-Scenario Survey ..............................................................................................82
LIST OF TABLES

Table 1. Overlap between frameworks for defining context. ...........................................34
Table 2. Definitions of usability. ..........................................................................................49
Table 3. Measurement scales for perceived usefulness (Davis, 1989). ............................52
Table 4. Three contextual scenarios used for the CONTEXT variable. ............................70
Table 5. Differences between contextual scenarios. ...........................................................71
Table 6. Items for measuring USEFULNESS. ..................................................................73
Table 7. Items for measuring USABILITY. .......................................................................74
Table 8. Items for measuring AESTHETICS. .................................................................74
Table 9. Items for measuring ENJOYMENT. ....................................................................75
Table 10. Item for measuring GOODNESS. .....................................................................76
Table 11. Summary of variables used in the study. .........................................................76
Table 12. Goodness of fit indices for confirmatory factor analysis. .................................84
Table 13. Sub-set of modification indices for the full model (Model 1). .........................87
Table 14. Measures and acceptance criteria for convergent validity. ............................91
Table 15. Convergent validity of the revised model (Model 2) ........................................92
Table 16. Measures and acceptance criteria for discriminant validity. ............................93
Table 17. Factor loadings of measurement items on the four constructs. .......................94
Table 18. Squared multiple correlations for each construct. ............................................95
Table 19. Correlations between the four dependent variables. .........................................96
Table 20. Possible covariance structures for the repeated measurements. ....................103
Table 21. Descriptive statistics for USEFULNESS. .......................................................109
Table 22. Summary of variables in initial marginal model for USEFULNESS. ..............111
Table 23. Likelihood ratio tests of covariance structures for USEFULNESS
Table 24. Final marginal model for USEFULNESS.
Table 25. Coefficient estimates for the final marginal model for USEFULNESS.
Table 26. Estimated correlation matrix for ratings of USEFULNESS.
Table 27. Descriptive statistics for GOODNESS.
Table 28. Summary of variables in initial marginal model for GOODNESS.
Table 29. Likelihood ratio tests of covariance structures for GOODNESS.
Table 30. Final marginal model for GOODNESS.
Table 31. Coefficient estimates for the final marginal model for GOODNESS.
Table 32. Estimated correlation matrix for ratings of GOODNESS.
Table 33. Overview of codes and themes from participant definitions of usefulness.
Table 34. Comparing dimensions of usefulness with usability measures.
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>History of evaluation in HCI.</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Four elements of context: user, task, system, environment (Shackel, 1991)</td>
<td>33</td>
</tr>
<tr>
<td>3</td>
<td>The structure of an activity (Kuutti, 1995, p. 28)</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>The Technology Acceptance Model (Davis, Bagozzi, &amp; Warsaw, 1989)</td>
<td>51</td>
</tr>
<tr>
<td>5</td>
<td>Usefulness/Ease of Use Grid (Keil, Beranek, &amp; Konsynski, 1995, p. 79)</td>
<td>55</td>
</tr>
<tr>
<td>6</td>
<td>Screenshots of ipl2, RefSeek, and Awesome Library.</td>
<td>69</td>
</tr>
<tr>
<td>7</td>
<td>Sets used to determine treatment combinations for each subject.</td>
<td>79</td>
</tr>
<tr>
<td>8</td>
<td>Hypothesized factor structure for the full model (Model 1).</td>
<td>85</td>
</tr>
<tr>
<td>9</td>
<td>Hypothesized factor structure for the revised model (Model 2).</td>
<td>89</td>
</tr>
<tr>
<td>10</td>
<td>Mean ratings of USEFULNESS by CONTEXT and SYSTEM.</td>
<td>110</td>
</tr>
<tr>
<td>11</td>
<td>Visual representation of the USEFULNESS marginal model.</td>
<td>115</td>
</tr>
<tr>
<td>12</td>
<td>Mean ratings of GOODNESS by CONTEXT and SYSTEM.</td>
<td>117</td>
</tr>
<tr>
<td>13</td>
<td>Visual representation of the GOODNESS marginal model.</td>
<td>123</td>
</tr>
<tr>
<td>14</td>
<td>Four dimensions of usefulness identified by participants.</td>
<td>137</td>
</tr>
</tbody>
</table>
ABSTRACT
Understanding Usefulness in Human-Computer Interaction to Enhance User Experience Evaluation
Craig Matthew MacDonald
Michael E. Atwood, Ph.D.

The concept of usefulness has implicitly played a pivotal role in evaluation research, but the meaning of usefulness has changed over time from system reliability to user performance and learnability/ease of use for non-experts. Despite massive technical and social changes, usability remains the “gold standard” for system evaluation. However, as user experience (UX) emerges as the dominant paradigm in HCI, it is necessary to consider whether usability is sufficient and if the meaning of usefulness needs to be updated to reflect the complexity of modern interactive computing experiences.

This dissertation describes the results of a repeated measures laboratory experiment to investigate the nature and meaning of usefulness and its relationship to common UX attributes: usability, aesthetics, and enjoyment. Quantitative and qualitative analyses showed that the usefulness of a system is shaped by the context in which it is used, that usability is a major element of usefulness, that usefulness has both pragmatic (e.g., usability, simplicity) and hedonic (e.g., aesthetics, pleasurable interactions) attributes, and that usefulness plays a pivotal role in defining users’ overall evaluation of a system (i.e., its goodness).

These results have several implications for evaluators of interactive systems: first, evaluators should be trained to look beyond usability and probe for issues related to usefulness; second, the scope of evaluation should be broadened to include both
pragmatic and hedonic elements; third, evaluators should vary evaluation contexts to simulate the complexity of real world interactive experiences. Future research will clarify and extend our understanding of usefulness by examining usefulness in other contexts, supplementing laboratory studies with naturalistic inquiries, and developing new evaluation methods that reflect the multi-faceted nature of usefulness.
1. INTRODUCTION

On a broad level, what people accomplish with computers has not changed dramatically over the past 30 years: people still make calculations with spreadsheets, communicate with friends, purchase goods, and search for information. Rather, the pervasiveness of the web and the blurring line between work and personal technology use (Bødker, 2006) has resulted in drastic changes to the context under which these activities occur. People may be doing the same (or similar) things with technology, but more people are doing more things in more places at more times and for more reasons than ever before. In response to this rapidly changing environment, researchers in Human-Computer Interaction (HCI) have developed new approaches to evaluating the user experience (UX) of systems, which aim to capture both the pragmatic and hedonic aspects of user-system interactions. While some of these methods have promise, they have not yet been widely adopted within the HCI community and usability evaluation remains the “gold standard” for evaluating interactive systems (Hornbæk, 2010). But, usability is not the only factor that determines the success or failure of a system; as computing pioneer Douglas Engelbart once reportedly remarked, “If ease of use was the only valid criterion, people would stick to tricycles and never try bicycles” (Beale, 2007, p. 21). Nevertheless, usability research has dominated HCI practice and research over the past 40 years. Although these efforts have yielded a great deal of insight about how to design systems that are easy to use and easy to learn, little time has been spent determining whether or not systems are useful.

The concept of usefulness is not new to the field of HCI. Designing useful systems has long been cited as one of the primary motivations for employing user-
centered design methods (e.g., Gould & Lewis, 1985; Bannon & Bødker, 1991; Nielsen, 1993), but there have been few attempts to define what “usefulness” means in HCI and there is no consensus about how to measure or evaluate system usefulness. Despite this dearth of research, it is reasonable to assume that the usefulness of a system is inextricably linked to the context in which it is being used. Although the nature and meaning of context has long been debated, several theories germane to the study of HCI–activity theory (Kuutti, 1995), situated action (Suchman, 1987), distributed cognition (Hutchins, 1995), and others–suggest that context is an emergent property of technology use and imply that a system is a tool that mediates, rather than defines, the context of human activity. From this perspective, relating usefulness with context seems obvious; yet, there is little empirical evidence describing the exact nature of the relationship.

Any attempt to define or measure system usefulness must also consider its usability, or ease of use, but little is known about how the two concepts are related. For example, it has been argued that a highly usable system may not be useful for anything while a highly useful system may still be used despite having numerous usability problems (Greenberg & Buxton, 2008). On the other hand, it is believed that severe usability problems may prevent users from realizing the potential usefulness of a system (Eason, 1984). Despite the obvious connection, there have been few empirical studies investigating the relationship between usability and usefulness and little is known about the exact nature of the relationship. Furthermore, recent research has been able to provide insight into the role of aesthetics and visual appeal in determining perceptions of usability (e.g., Tractinsky, 1997). More recently, researchers have developed frameworks for understanding the affective and emotional factors of using interactive systems (e.g.,
Forlizzi & Battarbee, 2004). These efforts have been successful in making designers aware that the visual appearance of a system and users’ affective states can influence their perceptions of usability, but it is unclear whether and how these factors may influence perceptions of usefulness.

In summary, the concept of usefulness has long been identified as a critical goal of employing user-centered design methods, but the construct of usefulness has been largely ignored in HCI research. It is widely believed that the usefulness of a system depends on the context in which it is being used, but there have been few attempts to verify this relationship. It is also expected that usefulness is closely related to usability and other system attributes (e.g., aesthetics and enjoyment) but there is little empirical evidence describing these relationships. As the field of HCI continues to transition from a narrow, task-centered emphasis on usability to a broader, more holistic focus on UX, it is necessary to consider the meaning of usefulness of systems in modern computing contexts and gain a firm understanding of how usefulness relates to the critical UX aspects of usability, aesthetics, and enjoyment. Thus, this research was guided by the following question: what does usefulness mean for modern interactive computer systems?

The following research questions were considered:

**RQ1:** How and to what extent are perceptions of usefulness shaped by the context—the user, task, and environment—in which the system is used?

**RQ2:** How and to what extent do perceptions of usability, aesthetics, and enjoyment influence perceptions of usefulness?

**RQ3:** To what extent do perceptions of usefulness affect users’ overall evaluation of a system?
**RQ4:** What criteria do users identify when defining or describing a useful system?

As one of the earliest and most ardent proponents of a user-oriented perspective to computing (Gaines, 1984), the practice and research of HCI has yielded valuable insight into how to design highly usable—and potentially useful—interactive systems. But what has been missing is knowledge of how to bridge the gap from “potentially” useful systems to “actually” useful systems. The first step in this direction is examining the construct of usefulness and its relationship to context and other major attributes of UX. A thorough understanding of usefulness may then be used to inform the development of more agile, flexible, and effective UX evaluation methods.

The remainder of this dissertation is organized as follows. In Chapter 2, the history of system evaluation is reviewed to show that usefulness has always been a critical goal of system evaluation but the meaning of the term “usefulness” has changed over time. The research literature is then examined to: 1) determine how others have defined usefulness and propose a working definition; 2) provide a framework for understanding context and its potential relationship to usefulness; and 3) examine the potential relationships between usefulness and usability, aesthetics, and enjoyment. Chapter 3 presents the research methods by describing a repeated measures laboratory experiment designed to address the research questions. In the experiment, 36 participants interacted with three information portal websites under three different contextual scenarios and were asked to rate each website in terms of usefulness, usability, aesthetics, enjoyment, and goodness (their overall evaluation). Participants were also asked how they would define the concept of usefulness. In Chapter 4, the results of the quantitative
and qualitative analyses are presented and it is shown that 1) usefulness is highly influenced by context, 2) usefulness and usability are closely related, 3) usefulness and aesthetics have a complex relationship, 4) usefulness and enjoyment are unrelated, and 5) users define usefulness in terms of a system’s fit to context, its simplicity and ease of use, its visual attractiveness, and its provision of pleasurable interactions. These results are further examined, discussed, and interpreted in Chapter 5. Chapter 6 concludes the research with implications for HCI research and practice and plans for future work in this area.
2. LITERATURE REVIEW

This chapter presents a literature review covering several areas of Human-Computer Interaction (HCI) research, including: system evaluation, user experience (UX), usefulness, context, usability, aesthetics, and enjoyment. First, the history of evaluation research in HCI will be discussed to determine how the concept of usefulness has evolved over time. Second, the construct of usefulness will be explored by examining common definitions of the term usefulness in HCI and a working definition of the term will be proposed. Third, perspectives on the nature and meaning of context will be explored and a simple framework for understanding context will be presented. Fourth, the close connection between usefulness and usability will be analyzed. Finally, research on aesthetics and enjoyment will be discussed in terms of how affective aspects of UX may influence the concept of usefulness.

2.1 History of Evaluation in HCI

In their review of the history of evaluation in HCI, Kaye and Sengers (2007) adapted the work of Grudin (2005) to identify four phases of evaluation by focusing on 1) who was doing the evaluation and 2) what role users played. This literature review expands on these four phases by also considering how changes in technology influenced the choice of evaluation methods. It will also identify a fifth phase to address the challenges of “third wave” computing (Bødker, 2006). The five phases will be referred to as the Reliability Phase (1950s and before), the System Performance Phase (1950s-1960s), the User Performance Phase (1960s-1970s), the Usability Phase (1980s-2000s), and the User Experience (UX) Phase (2000s-Present). These phases are presented graphically in Figure 1. Although the phases are presented as discrete time periods, the
years are meant to provide historical context and should not be taken literally. In reality, there was a substantial overlap between the phases as evaluators transitioned from one method to the next.
Figure 1. History of evaluation in HCI.
2.1.1 The Reliability Phase (1950s and before)

In the Reliability Phase (1950s and before), computer systems were incredibly large and complex machines whose operation often required the manipulation of switches, lights and plugs. Because operating these machines required a high degree of technical expertise, users (and evaluators) were primarily engineers. Since computer systems during this period were generally used to perform complex calculations on large quantities of data, evaluation typically focused on “reliability” in terms of how long a computer would function without failure (Kaye & Sengers, 2007). For example, in an evaluation of a “large-scale general purpose digital computer” Pollard (1951) focused on methods of minimizing system fault time and stressed that “availability” was the most critical element of system performance. As Kaye and Sengers (2007) noted, the emphasis on reliability is because the “limiting factor in determining how well a computer worked was how long it would do so” (p. 3). In summary, since computers were primarily expected to perform complex mathematical calculations over long periods of time, “the usefulness of a computer… [was] entirely judged by its reliability in operation and the ease of maintenance” (Pollard, 1951, p. 62).

2.1.2 The System Performance Phase (1950s-1960s)

In the System Performance Phase (1950s-1960s), computer systems were still relatively large, but they were beginning to shrink considerably compared to their predecessors. Instead of manipulating plugs and switches, computer systems could now be used with magnetic tape, punch cards, light guns, and, eventually, keyboards. These improvements helped lead to the development of early programming languages like FORTRAN and COBOL, which meant that users (and evaluators) shifted from engineers
to programmers and computer scientists. While the focus was still largely on reliability, the meaning of “reliability” shifted away from how long a system would perform to how quickly a system would perform (Kaye & Sengers, 2007). Primarily motivated by the economic impact of using a computer to perform complex tasks, the focus of evaluation during this period was on system performance, particularly processing speed. In evaluations of the UNIVAC system, for example, Alexander and Elbourn (1953) describe acceptance tests based on “estimates of the minimum performance [time] which would be necessary for the machine to be economically competitive with the conventional tabulating-card methods of doing Census work” (p. 58). As a result, the tests were designed to evaluate how long it would take the system to process large amounts of data with minimal down time and minimal errors (McPherson & Alexander, 1951). As another example, Israel (1957) described the use of simulation techniques to test and evaluate computer systems for “safety and adequacy of performance” (p. 354). Kaye and Sengers (2007) also found references to throughput, turnaround, and availability as measures of performance, with a specific focus on “how fast [the system] runs” (p. 4). To conclude, the meaning of usefulness during this time period shifted from reliability in terms of downtime/availability to system performance in terms of processing speed and processing errors.

2.1.3 The User Performance Phase (1960s-1970s)

In the User Performance Phase (1960s-1970s), evaluation shifted from a primarily systems-based focus to a more user-centered perspective, mostly due to a substantial change in computing use brought about by time-sharing systems. By the late 1960s, the large-scale batch-processing machines were slowly getting replaced by more expensive,
but supposedly more efficient, time-sharing machines (Grant & Sackman, 1967). Advocates of time-sharing systems claimed that they allowed for faster program development and saved programmer time, while critics argued that the systems were less efficient since less computational power was used on productive programming tasks (i.e., programmers wasted more time). To address these issues, Grant and Sackman (1967) presented an evaluation of programmer performance in both on-line (time-sharing) and off-line (closed-shop) conditions, describing their study as “a pioneering effort in the collection of performance data for computer programmers under controlled conditions” (p. 33). The authors measured performance using man-hours – the amount of time spent debugging by each individual user – and found that, when controlling for programmer experience, programmers were significantly more efficient in the on-line condition compared to the off-line condition (Grant & Sackman, 1967).

As time-sharing systems grew in popularity, it became clear that, for the first time, people were using computers for non-programming tasks (e.g., text editing). The introduction of “non-specialists” into the equation was itself a major shift because it forced evaluators to be less interested in evaluating the speed of the computer and more interested in the speed of the user. Consequently, evaluators began to focus on “the speed of execution of… tasks… rather than the speed of the calculations that enable that execution” (Kaye & Sengers, 2007, p. 5). It was also during this time that the field of HCI emerged as a discipline. Several conferences began holding special sessions on human factors, and the first issue of the International Journal of Man-Machine Studies was published in 1969 as the first publication entirely devoted to human factors research (Gaines, 1984). This increased emphasis on user performance is perhaps best captured by
Sime, Green, and Guest (1973), who noted that “…quite apart from the power of a system, its efficiency or its cost, it is becoming increasingly important that it should be amenable to the inexpert” (p. 105).

Due to this interest in studying users, system evaluation methods were typically laboratory-based experiments. Since use contexts were almost exclusively limited to work-place applications, these methods were mainly focused on enhancing worker productivity through the use of performance-based metrics. In one of the first studies of its kind, English, Engelbart, and Berman (1967) found that a newly invented input device called a “mouse” was more effective for selecting bits of text on a computer screen than a light pen, joystick, Grafacon, or knee control device. To reach this conclusion, the authors simulated “conditions similar to those that the user would encounter when actually working on-line” (p. 7) by asking both novice and expert users to complete a series of point-and-select tasks, measuring their speed (access time and motion time), ease of learning, error rate, accuracy, and satisfaction with each device. Later, in what Gaines (1984) referred to as “the beginnings of an experimental psychology interest in HCI” (p. 159), Sime et al. (1973) measured task completion rate, task completion time, and number of errors to evaluate the performance of novice programmers using two different programming languages.

It was also during this time that researchers began publishing principles for effective system design. Some were based on empirical work (e.g., Dzida, Herda, & Itzfeldt, 1978), but the majority of researchers used anecdotal evidence culled from personal experiences. According to Gaines (1984), “the first attempt to tabulate some user engineering principles for the design of interactive systems” (p. 158) was presented
by Hansen (1971), who described lessons he had learned while designing the Emily text editing system. Reflecting on his experience, Hansen implored designers of interactive systems to “know thy user” and to focus on what users need to accomplish by minimizing memorization, optimizing operations, and engineering for errors. Conversely, Dzida et al. (1978) relied on survey data and factor analysis to develop a different set of design principles, finding that the five major factors of user-perceived quality of interactive systems were self-descriptiveness, user control, ease of learning, problem adequate usability, and correspondence with user expectations.

Understanding computer users is a consistent theme during this period, but there seemed to be disagreement about what exactly what the term “user” meant. On one end of the spectrum, Wasserman (1973) presented principles for designing “idiot-proof” systems, based on the axiom that “any error that can be made will be made” (p. m35). Although Wasserman’s terminology may be somewhat insensitive by today’s standards, the principles he provided – providing actions for every possible type of user input, minimizing the amount of learning, providing explicit diagnostics, providing program short-cuts, and allowing users to express the same message in multiple ways – stress the importance of error prevention, which is still a consistent theme in HCI today. At the opposite end of the spectrum is Jones (1978), who argued that very few problems people encounter when using computers result from unsophisticated software or hardware; rather, they are mostly “people problems” (p. 198) which can be avoided by developing a better understanding of how people interact with computers. Jones developed four principles of “effective man-computer dialogue” (expectations, implication, experimentation, and motivation), emphasizing that a “user” is a human being with
thoughts and feelings and a “system” is simply a tool that could be used to solve problems. Finally, Nickerson (1981) expanded the notion of user-system interaction by looking at why people may not use a computer system even if they might benefit from it. The problems he identified included accessibility and availability, work-session interrupts or crashes, inadequate training and user aids, unclear or incomplete documentation, complex command languages that are difficult to learn, inaccurate user conceptualizations of systems, and miscellaneous other personal or job-related factors (e.g., resistance to change, fear of dehumanization). Although researchers were still grappling with the complexities of designing for people, the publication of user-centered design principles combined with the widespread use of user performance measures indicated a substantial shift in system evaluation from a system-based to a user-centered perspective. Consequently, the meaning of usefulness shifted from system performance in terms of processing speed and processing errors to user performance in terms of user speed and user errors.

2.1.4 The Usability Phase (1980s-2000s)

In the Usability Phase (1980s-2000s), drastic reductions in the size and cost of personal computing workstations greatly increased their appeal to non-programmers. According to Lindgaard and Parush (2008), this appeal was not a direct result of the invention of WIMP (windows, icons, menus, pointers) in 1973 or the launch of Xerox Star in 1981; rather, the personal computer did not gain commercial success until 1984, when the Apple Macintosh was released with the first direct manipulation spreadsheet program, Visicalc. The sudden demand in office productivity software led to an increase of in-house software development projects aimed at streamlining existing business
practices. These software systems relied heavily on user manuals which were usually out-of-date or inaccurate, and “in this climate, the practice of usability began to prosper” (Lindgaard & Parush, 2008, p. 223). The subsequent increase in the number of novice users challenged designers to develop systems that could be used by both experts and novices with minimal training and support, thus HCI and usability professionals developed methods for evaluating the “ease of use” of computer systems because “if a system is not easy to learn, it will not be used…[and] if a system is cumbersome to use it will either be circumvented or it will be used in its own inefficient way” (Good, 1982, p. 143). Accordingly, evaluation efforts were expanded to encompass aspects of learnability and ease of use in addition to speed and efficiency in user performance.

The widespread interest in usability throughout the field of HCI may have been largely driven by Gould and Lewis (1985), who proposed three design principles of designing for usability: 1) early focus on users and tasks, 2) empirical measurement with actual users and a prototype system, and 3) iterative design and development. To determine whether these principles were used in practice, they surveyed almost 450 professional system planners, designers, programmers and developers to see how many of the principles were being followed. They found that, despite the apparent obviousness of these principles, only 16% of those surveyed mentioned all three while 26% mentioned none of the three. They argued that although the process may be time-consuming and expensive, failing to address critical usability problems will be more time-consuming and expensive in the long run (Gould & Lewis, 1985).

Researchers during this time period began developing standardized methods to evaluate usability, with an early emphasis on evaluating the ease of use of text editors. In
a broad review of this research, Embley and Nagy (1981) observed that a variety of methods had been used to study text editors (introspection, field studies, formal analyses, controlled experiments, and psychological models), but most research activity focused on measuring performance time in “symbol manipulation in tasks of very limited scope where really significant improvements over the best of present-day editors appears unlikely” (p. 63). To remedy this apparent deficiency, Roberts and Moran (1982) developed a comprehensive method for evaluating text editors by measuring how long it took experts to perform tasks, how many errors experts encountered, how long it took novices to learn basic tasks, and the number of functions the application offered compared to its competitors. They tested this method on eight different text editors and found that no text editor was superior on all four dimensions, concluding that users must make tradeoffs depending on the situation and the application.

It was also during this time period that Card, Moran, and Newell’s (1980; 1983) GOMS models (goals, operators, methods, and selection) were being used to develop models of human performance. While these models were considered “cumbersome and time consuming to build” (Lindgaard & Parush, 2008, p. 224), they nevertheless represent one of the first attempts to empirically address the concept of usability. At this time, researchers also began formalizing the process of user testing with “think aloud” protocols (Lewis and Mack, 1982), which soon became one of the most popular methods of evaluating system usability. In the 1990s, researchers began searching for alternative methods to the often time-consuming process of user testing. Methods such as heuristic evaluation (Nielsen & Molich, 1990) and cognitive walkthrough (Lewis, et al., 1990) were developed as so-called “discount” or “analytic” methods because they aimed to
replace empirical user observation with usability expertise and knowledge. Regardless of the method, it is clear during this time period that a useful computer system was a usable system; that is, it was easy to learn and easy to use by novices.

2.1.5 The User Experience (UX) Phase (2000s-Present)

Finally, in the User Experience (UX) Phase (2000s-Present), personal computing, social computing, mobile computing, and cloud computing have drastically altered the contexts in which people use interactive computer systems (Bødker, 2006). These developments have led to the rise of UX as “a new paradigm” for designing and evaluating modern interactive systems (Bargas-Avila & Hornbæk, 2011). The concept of UX tends to focus on the emotional factors that influence, and sometimes supersede, usability as a determinant of system quality. While there is no consensus about a precise operational definition of the term, there is broad agreement that UX research “highlights non-utilitarian aspects of…interactions, shifting the focus to user affect, sensation, and the meaning as well as value of such interactions in everyday life” (Law et al., 2009, p. 719). Thus, UX research takes a much broader perspective than usability research because it “encompasses all aspects of interacting with a product” (Hassenzahl, 2003, p. 41) and UX researchers attempt to understand and assess all stages of the interaction in order to fully capture how users interact with a system (Forlizzi & Battarbee, 2004; Hassenzahl 2003; Law et al., 2009). Rather than the narrow, task-based focus of usability, UX researchers take a holistic view of the entire user-system interaction by going beyond the tasks users perform with technology and focusing on the provision of positive emotional outcomes from using technology (Hassenzahl & Tractinsky, 2006; Lindgaard & Parush, 2008; Frandsen-Thorlacius et al., 2009). As a result, UX is subjective, varies
between individuals and between situations, and changes over time. To put it simply, the UX perspective forces designers to focus on creating a positive experience for users rather than on preventing usability problems, or “designing for pleasure rather than for absence of pain” (Hassenzahl & Tractinsky, 2006, p. 95). Therefore, it is likely that the meaning of usefulness needs to be modified once again to address both the practical and affective aspects of system use.

Despite its popularity, there is still no shared conceptual framework for understanding UX. In a survey of 275 individuals from the UX community, Law et al. (2009) found that although the participants “understand the notion of user experience very differently” (p. 726), a majority of them agreed that UX was a dynamic, subjective, context-dependent property that can be predicted and, ultimately, addressed via the design process. Consequently, many different models of UX have been proposed. Hassenzahl (2003) argued that a product has pragmatic attributes, which describe a product’s functionality and are used to achieve behavioral goals, and hedonic attributes, which refer to a product’s ability to evoke feelings of pleasure, allow for self-expression, and provoke memories. While not explicitly a model of UX, Norman (2004) advocated for “emotional design” that consists of three levels corresponding to different types of experience: visceral, behavioral, and reflective. The visceral level is determined by the aesthetic appeal of a product and how people initially perceive and react to it, the behavioral level is determined by the function a product performs and how well it serves its intended purpose, and the reflective level is about emotion and how people construct the meaning of products they use (Norman, 2004). Forlizzi and Battarbee (2004) described three types of user interactions: fluent interactions, which are automatic,
invisible and allow users to focus on outcomes, *cognitive* interactions, which cause some type of change in users’ knowledge or context of use, and *expressive* interactions, which allow users to modify or adapt products to fit their own personal needs. Taking it one step further, Karapanos et al. (2009) developed an initial framework for “user experience over time” consisting of three phases in users’ adoption of the iPhone: orientation (familiarity), incorporation (functional dependency), and identification (emotional attachment). They concluded that “the overall value, or the *goodness* of a product is contextually dependent... [so] the focus of CHI practice should expand from the study of early interactions to the study of prolonged experiences, understanding how a product becomes meaningful in a person’s life” (p. 737). While each of these models provides a valuable perspective on understanding the complex nature of UX, research in this area is still in its early stages.

One challenge facing UX research is that emotions are subjective and highly influenced by a multitude of external factors (Tractinsky, Katz, & Ikar, 2000; van Shaik & Ling, 2009). A second challenge is that the widespread use of cloud computing and mobile devices creates use contexts which are increasingly complex and difficult to predict beforehand. To address these challenges, several evaluation methods have been proposed. Cockton (2006) proposed that designers concentrate on providing some type of value or worth to users by focusing on “what endures beyond interaction” (p. 166) rather than what occurs during the interaction itself. He proposed a method called Worth Mapping, which assesses “the impact of user experience and performance on [users’] achievement of intended worth” (p. 172). Bertelsen (2004) described an approach called the Activity Walkthrough which blends activity theory with the cognitive walkthrough.
The Living Laboratory method (e.g., Abowd, 1999) allows designers the opportunity to observe users interacting with new technology in a pseudo-realistic testing environment. Each of these methods is aimed at capturing UX in the complex use contexts typical of modern computing, but they have not yet been widely adopted within the HCI community. An exception may be the living laboratory approach which has been shown to be successful in evaluating ubiquitous computing technologies (e.g., Intille, et al., 2005); however, an effective living laboratory requires a great deal of time, effort and resources, limiting its appeal to the HCI community.

Although it is common to associate non-instrumental/hedonic goals with UX and instrumental/pragmatic goals with usability, it is becoming increasingly clear that the two concepts are interwoven (Bargas-Avila & Hornbæk, 2011). Therefore, despite the popularity of UX as a concept, the usability evaluation methods developed in the 1980s and 1990s—user testing, heuristic evaluation, and cognitive walkthrough—are still considered the “gold standard” for evaluation today (Hollingsed & Novick, 2007; Hornbæk, 2010). In this environment, it is questionable whether these methods, which were initially developed to evaluate ease of learning for novice users, are still appropriate for evaluating the multifaceted construct of UX. Additionally, the uncertainty and lack of consensus surrounding UX means that the meaning of usefulness in this complex computing environment is unclear.

2.1.6 Summary

The previous sections detailed the history of system evaluation in HCI from the early 1950s, where the focus was on reliability, to the present day, where the focus is on user experience (UX). The examples cited from each phase suggest that measuring
system usefulness was always one of the primary goals of evaluation, but the meaning of usefulness has changed over time. In the 1940s and 1950s, a computer system was only considered useful if it could process large amounts of data reliably (i.e., with minimal downtime and errors); in the 1960s and 1970s, a useful computer system was one that allowed its users to perform tasks more quickly and efficiently than other methods; in the 1980s, due to the sudden influx of novice users, computer systems were considered useful if they were easy to learn and easy to use. Today, due to the increasing complexity, diversity, and unpredictability of use contexts, usefulness is likely related to the multifaceted (but ill-defined) construct of UX but the exact meaning of usefulness remains unclear.

2.2 Defining Usefulness

Designing useful systems has long been cited as one of the primary goals of user-centered design. As an early example, Dreyfuss (1955) listed Utility and Safety as the first point of his Five Point Formula for industrial design. In their seminal work, Gould and Lewis (1985) stated that “any system designed for people to use should be easy to learn (and remember), useful…, and be easy and pleasant to use” (p. 300). Bannon and Bødker (1991) describe their view of HCI as “a design discipline that has as its goal the provision of more usable (and hopefully useful) artifacts” (p. 4). But despite these authors’ insistence that usefulness is one of the core components of design, there is no standard, agreed-upon definition of usefulness in HCI. Even Grudin’s (1992) paper discussing the difference between “utility” and “usability” does not contain any formal definition. This section will explore how the term usefulness has been defined and used in HCI and propose a working definition.
While some HCI researchers make a clear distinction between “utility” and “usefulness” (e.g., Nielsen, 1993) there seems to be substantial overlap in the way both terms are used and defined in the literature. Therefore, this research will consider definitions of both “utility” and “usefulness” in order to provide a more complete understanding of the concept. A review of the literature revealed 28 different definitions or uses of the terms “usefulness” or “utility,” each of which covered one or more of the following:

1. The **functions** provided by the system; e.g., “the functions people really need in their work” (Gould & Lewis, 1985),

2. The **tasks** users are trying to complete; e.g., “the role of the technology in accomplishing the user’s relevant tasks” (Maryniak-Nelson & Caldwell, 1992),

3. The **goals** users are trying to achieve; e.g., “whether system can be used to achieve some goal” (Nielsen, 1993), and

4. The **context** in which the system is being used; e.g., “how technology can fit into users' actual social and material environments” (Nardi, 1996b).

These four elements—functions, tasks, goals, and use context—represent different levels of usefulness in terms of scope and breadth, with functionality referring to usefulness in the narrowest sense and use context in the broadest. Together, the four elements provide a framework for an initial working definition of usefulness as “the extent to which a system’s functions allow users to complete a set of tasks and achieve specific goals in a particular context of use.” Each of these elements will be discussed in the following sections.
2.2.1 Functions

Many researchers used the term *usefulness* to refer solely to system functionality. Eason (1984) used utility as a cover term for the amount of functionality the system is able to provide. Shackel (1991) used the term utility to describe whether the system will “do what is needed functionally” (p. 22). Grudin (1992) made a distinction between utility and usability by describing the difference between a system’s functionality and its interface\(^1\). Tractinsky (1997) noted simply that utility is a system attribute that describes its functionality. Maryniak-Nelson and Caldwell (1992) noted that utility “refers to the function that the product performs” (p. 879). Nielsen (1993) defined utility as “the question of whether the functionality of the system in principle can do what is needed” (p. 25). In his model of user experience, Hassenzahl (2003) identified utility as a pragmatic quality of systems use and defines it simply as “relevant functionality” (p. 12). Hartmann, Sutcliffe, and De Angeli (2007), in their study of user satisfaction with websites, used the term “services” to describe “the functions that aggregate into utility” (p. 388). Thus, at its most basic level, usefulness refers to the amount and type of functions provided by a system.

2.2.2 Tasks

The term *usefulness* has also been used to describe the tasks a system does (or should) support through the functions it provides. For example, although Hassenzahl (2003) used the term “utility” to refer to functionality, he also noted that it should only be considered “in relation to [the] potential tasks” (p. 12) users want or need to complete.

---

\(^1\) Grudin’s (1992) main argument was that utility and usability are rarely (if ever) considered at the same time, partly because the concepts belong to two different research areas: he argued that usability fell in the domain of HCI research (and the CHI conference) while utility fell in the domain of IS research (and the International Conference on Information Systems).
Similarly, Tractinsky (1997) noted that functional utility is “predominated by specific tasks” (p. 115) and Maryniak-Nelson and Caldwell (1992) argued that “utility refers to the role of the technology in accomplishing the user’s relevant tasks” (p. 879). Gould and Lewis (1985) argued that a useful system should “contain functions people really need in their work” (p. 300). In addition, although not using the term explicitly, Goodwin (1987) observed that a good system should “provide the functions a user needs to accomplish a task or set of tasks” (p. 229). Chen, German, and Rorissa (2009) defined usefulness as whether “users find the content/information useful for their needs and tasks” (Methodology section, Table 1). Dillon and Morris (1993) defined utility as “the technical capability of the tool to actually support tasks that the user wishes to perform” (p. 6). In their study of usability practitioners, Nørgaard and Hornbæk (2006) defined utility as the “tasks the system should support or whether tasks from the [usability] test were unrealistic with regard to how the user usually uses the system or would want to use the system” (p. 215). Borrowing from patent law, Hudspith (1997) observed that utility can be objectively measured by “how well an artifact accomplishes or facilitates the ‘task’, i.e. does it do the practical things it is supposed to do: does it ‘facilitate’ the user’s task and provide the right tools to do so?” (p. 448). Hartson (2003) discussed how the usefulness of systems is related to their “functional affordances”, or those design features that “help users accomplish work” (p. 323). Scholtz (2006) associated utility with impact measures in order to answer questions like “how do human information interactive systems change the work that users are doing? What affect [sic] do new information interaction systems have on work products?” (p. 512). The above definitions suggest that
a useful system must not just provide functions, but provide appropriate functions to help users complete their tasks.

2.2.3 User Goals

The term usefulness has also been used to describe the goals users wish to achieve using the system. This perspective recognizes that “usefulness refers to the overall usefulness of the product. Does it do what it is supposed to? ...Does it work?” (Dicks, 2002, p. 29). Seffah et al. (2006) defined usefulness as “whether a software product enables users to solve real problems in an acceptable way” (p. 169) and Nielsen (1993) defined usefulness as “the issue of whether the system can be used to achieve some desired goal” (p. 24). Similarly, Hong and Kim (2004) defined the utility of a website as whether or not it provides the appropriate functions to “help users accomplish their goal” (p. 339). McGrenere and Ho (2000) argued that “the usefulness of a design is determined by what the design affords ...and whether these affordances match the goals of the user and allow the necessary work to be accomplished” (p. 184). Olson and Moran (1996) argued that a system has utility if it is able to meet the needs of users and solve their problems. From these definitions, the implication is that a useful system must provide appropriate functions to help users complete their tasks and achieve their goals.

2.2.4 Use Context

At the fourth level, the term usefulness describes the context within which the user-system interaction takes place. One of the first to describe a contextual relationship with usefulness was Dreyfuss (1955), who listed Utility and Safety as the first point of his Five Point Formula for industrial design. Although Dreyfuss did not explicitly define the
word utility, he illustrated the concept using a description of what he believed was a well-designed vacuum cleaner:

To make an upright cleaner push and pull as comfortably as possible, we consider the length of the handle with relation to the specifications of Joe and Josephine and resolve the details of the handle grip to fit the hand... The height must be kept to a minimum, so that the cleaner will slide under furniture. A rubber bumper protects walls and furnishings. Suitable handgrips for carrying the cleaner from one place to another are necessary. The size must be such that the cleaner can be conveniently stored in minimum closet space. (p. 179)

From this quote, it is clear that Dreyfuss envisioned a useful vacuum cleaner as one that is appropriate for a number of different situations in which it will likely be used. Hansen (1971) made a similar point when he remarked that the most important element of successful design is not just “know[ing] thy user” but actually “tailoring the system to the user” (p. 572). Additionally, although Olson and Moran (1996) remarked that utility is about meeting users’ needs, they also noted that systems “must... meet the needs in the context of the work setting (fitting the organizational culture, the established communication patterns, and the incentive structure, among other things)” (p. 272). Furthermore, Karapanos et al. (2009) defined usefulness as how well a product supports users' daily activities, but they found that it was also closely related to the broader user experience; specifically, study participants identified the iPhone as a useful tool for “providing fast access to information”, “alleviating boredom in idle periods”, “capturing momentary information “, and “avoiding negative social situations.” These findings suggest that usefulness is “much broader than the functionality of the product, relating to the impact of the functionality in participants’ lives” (Karapanos et al., 2009, p. 737). The
connection between usefulness and context is perhaps best captured by Nardi (1996b), who noted that:

As we expand our horizons to think not only about usable systems but now useful systems, it is imperative that we have ways of finding out what would be useful... [in terms of] how technology can fit into users' actual social and material environments, the problems users have that technology can remedy, the applications that will promote creativity and enlightenment, and how we can design humane technology that ensures privacy and dignity. (p. 8-9)

These definitions of usefulness imply that the usefulness of a system cannot be separated from the context in which it is used. It should be noted that context is also assumed to be an important aspect of usability (ISO, 1997) and user experience (Bargas-Avila & Hornbæk, 2011). However, it seems from this analysis that context plays a particularly strong role in determining system usefulness. As a result, a working definition of usefulness is proposed as “the extent to which a system’s functions allow users to complete a set of tasks and achieve specific goals in a particular context of use.”

2.2.5 Summary

In HCI, use of the term “usefulness” has fallen into four broad categories: the functions a system provides, the tasks users need to complete, the goals users wish to achieve, and the context in which the system is used. A working definition of usefulness is proposed as “the extent to which a system’s functions allow users to complete a set of tasks and achieve specific goals in a particular context of use.”

2.3 Usefulness and Context

This section will explore the nature and meaning of context as it pertains to the study of HCI. First, it will be shown that there is no coherent knowledge base regarding usefulness and its connection to context by examining three disparate studies which
investigated the contextual aspect of usefulness. Second, a framework for understanding context will be presented, followed by a discussion of each individual element within the framework.

2.3.1 Studies of Usefulness and Context

Despite the stated importance of context in relation to usefulness, a literature review yielded just three studies (Abdelnour-Nocera, Dunckley, & Sharp, 2007; Hsi, 2004; Blandford, et al., 2008) that have explicitly recognized and attempted to define the contextual aspect of usefulness. While these studies did recognize and define context, each study employed different methods and defined context differently. Since there is no unifying framework for understanding the relationship between usefulness and context, interpreting the results of these studies is challenging, and their applicability to the field of HCI is unclear.

Abdelnour-Nocera et al. (2007) argued that usefulness is not a property of a system but rather is socially constructed in situ by users who are actually using it. Using this perspective, Abdelnour-Nocera et al. performed an 18-month ethnographic study in which they employed the concept of technological frames to describe how users developed perceptions about the usefulness of an enterprise resource planning (ERP) system. The authors identified four categories that define users’ technological frames and influence users’ perceptions of usefulness: goals, problems, elements of interpretation, and elements of practice. While these findings have clear implications for the design and implementation of ERP systems, there are two weaknesses in the study that limit its applicability to HCI research. First, these methods are designed to provide an in-depth investigation of one particular context rather than produce generalizable results that can
be applied to the field of HCI. An 18-month ethnographic study coupled with interviews, observations, and document analysis can yield remarkable insight about one specific system in one specific context, but it is not a realistic solution for a majority of HCI researchers and practitioners seeking to understand the usefulness of the systems they design. Second, the definition of usefulness employed in this study implies that the “technological usefulness of software artifacts remains unknown, or known only indirectly until they are in use” (p.155). The authors’ assumption that usefulness cannot be assessed prior to system deployment provides little guidance on how to design useful systems from the start.

In another study, Hsi (2004) defined usefulness as “the extent to which an application succeeds in assisting a set of users to achieve a set of goals, relative to the amount of effort required to engage those features” (p. 28). Hsi measured usefulness by determining how well the ontology of a system mapped to the domain of the user and then described three case studies to demonstrate the method, which consisted primarily of an “ontological excavation” where the interactive components of an interface are visually represented using a “morphological map.” These core components were then measured by their frequency of use to create use case silhouettes; if the core concepts were used often, then the interface was said to be useful. While this approach is a unique way of evaluating the “conceptual fitness” of a system, the connection between conceptual fitness and usefulness is unclear; the implicit assumption of Hsi’s approach is that usefulness is simply a tradeoff between functionality and effort, which ignores the other contextual factors (physical, socio-cultural, organizational, etc.) that may also play a role in determining system usefulness. Like the study by Abdelnour-Nocera et al. (2007),
Hsi’s approach may be helpful for improving an existing system but offers little insight into the process of designing a useful system from the start.

In the third and final study connecting usefulness and context, Blandford et al. (2008) described Concept-based Analysis of Surface and Structural Misfits (CASSM), a method for evaluating system utility where usefulness is a measure of “mis-fit” between systems and users. The CASSM method focuses on the concepts that are both embedded within the system and commonly understood by users who wish to interact with the system. Once these concepts are identified, an analyst identifies the actions and relationships relevant to each concept; by comparing the system concepts to the user concepts, analysts can identify both surface and structural “misfits.” The authors demonstrated the usability and usefulness of the CASSM method with three case studies: a robotic arm, a digital library, and a drawing tool. Informal evaluations of the method suggest that while CASSM is a helpful tool in many cases for identifying conceptual misfit, it is somewhat difficult to learn for novice analysts. In addition, the CASSM approach is still unclear about whether identifying conceptual mis-fit improves the usefulness of a system because it only captures elements of context that are directly related to the user and the system and not the overarching context in which the user-system interaction is embedded.

As illustrated from these three studies, research into the relationship between usefulness and context has been severely limited by a lack of a unifying framework. Specifically, the impact of each study is limited because each has adopted different definitions of both usefulness and context. Blandford et al. (2008) described a method to

---

2 For instance, it is more useful for a travel website to work in terms of journeys between places rather than flights between airports because a user’s goal in booking a flight is not simply to fly to an airport but to travel to a location (Blandford, et al. 2008).
identify mis-fits between the system and its users but did not explain how mis-fits relate to system usefulness; Hsi (2004) defined usefulness as a tradeoff between effort and benefit and developed a method to evaluate a system’s conceptual fitness, but ignored other elements of context and did not provide any guidance for how designers can build useful systems from the start; Abdelnour-Nocera et al. (2007) found that perceptions of usefulness are created *in situ* by users, but the method they employed – an 18-month ethnographic study – is not scalable to a real world setting and can only be used to improve the usefulness of existing systems. Although these studies have provided sound insight into one particular aspect of the relationship between usefulness and context, the lack of a unifying framework for understanding the two constructs has limited their impact.

### 2.3.2 Defining and Understanding Context

In HCI, there are multiple, varied opinions about what “context” is. Some researchers define context in terms of users’ physical location (e.g., Schilit & Theimer, 1994) while others view context as a complex interactional property of objects and actions (e.g., Dourish, 2004), which may or may not include socio-cultural or historical factors (e.g., Räsänen & Nyce, 2006). Acknowledging this apparent complexity, Connolly, Chamberlain, and Phillips (2008) defined context more broadly as “an analytical construct applying to those phenomena that: a. surround a given focal event, b. are relevant to that focal event, and c. are liable to enter into dynamic interplay with that focal event” (p. 48). Dey, Abowd, and Salber (2001) take a similarly broad approach, defining context as “any information that can be used to characterize the situation of

---

3 This debate is not exclusive to HCI. For example, after an extensive review of the Library and Information Science (LIS) literature, Dervin (1997) remarked that “there is no term more often used, less often defined, and when defined defined so variously as context” (p. 14).
entities (i.e., whether a person, place, or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves” (p. 106). But regardless of how it is defined, there seems to be agreement that context is a dynamic, ambiguous, amorphous concept that is ill-defined, at best, in a way that is useful to HCI (Greenberg, 2001; Kirsh, 2001; Bødker, 2006).

Despite this ambiguity, there have been several attempts to develop frameworks for understanding context. For instance, Connolly et al. (2008) drew from linguistics and developed a hierarchical framework of context, starting with cognitive context and external context. They broke cognitive context into two parts: mental context is how a user represents context in his or her mind while informatic context is how the system represents context. They split external context into discoursal context, or those elements directly related to the user-system interaction, and situational context, or those elements outside of the immediate discourse but still relevant to it. From the perspective of mobile and ubiquitous computing, Bradley and Dunlop (2005) identified five dimensions of context relevant to context-aware applications: task, physical, social, temporal, and cognitive. These elements take a system-centered approach to understanding context. For instance, physical context referred to the location, orientation, and state of the user and surrounding physical objects, while social context referred to the presence and behavior of surrounding people, etc. More recently, Wigелius and Vääätäjä (2009) identified five relevant dimensions: infrastructural context (technologies, network connections, devices), spatial context (place, temperature, noise, lighting, furniture), temporal context (schedules, deadlines, pace of work, time of day/week, regularity), social context (work community, persons present, culture), and task context (interaction when using the
system, entity and goal of tasks). A much simpler and straightforward framework for context was proposed by Shackel (1991) and adopted by Maguire (2001) and Carroll (2000). This framework contained four main elements: the user, the task, the tool, and the environment (see Figure 2). This framework may be considered a simplified version of the “activity triangle” which is associated with activity theory (Kuutti, 1995) and will be discussed in upcoming sections.

![Figure 2. Four elements of context: user, task, system, environment (Shackel, 1991)](image)

All of these frameworks described above present different perspectives on understanding context, but there is substantial overlap between them (see Table 1). Specifically, the four main elements of Shackel’s model–user, task, tool, and environment–are well represented across all four frameworks. In addition, this framework is nearly identical to the definition of context provided by the International Standards Organization (ISO), who defined context of use as “users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used.” Each of these elements will be discussed in subsequent sections.
Table 1. Overlap between frameworks for defining context.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>Mental context</td>
<td>Cognitive</td>
<td>Task context</td>
</tr>
<tr>
<td>Task</td>
<td>Discoursal context</td>
<td>Task</td>
<td>Infrastructural context</td>
</tr>
<tr>
<td>Tool</td>
<td>Informatic context</td>
<td>Physical</td>
<td>Spatial context</td>
</tr>
<tr>
<td>Environment</td>
<td>Situational context</td>
<td>Temporal</td>
<td>Temporal context</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social</td>
<td>Social Context</td>
</tr>
</tbody>
</table>

2.3.3 The Elements of Context

This section will explain the main elements of context identified by the framework presented above: the user, the task, and the environment (both physical and social/cultural).

2.3.3.1 The User

The fact that the user is a critical element of context does not seem to be a matter of debate. Rather, the challenge for designers has always been determining which characteristics of users are most relevant to the design of interactive systems. In the early days of HCI (i.e., the “first wave”), the predominant view was that designers only needed to understand the basic cognitive processing abilities of the people who will be using the system (Bannon, 1990; Kuutti, 1995). From the traditional cognitive science perspective, Hutchins (1995) remarked that intelligence was viewed as inside the “inside/outside boundary” (p. 355), and users were seen merely as elements of a “mechanized formal symbol processing system” (p.365). There are, however, many other relevant dimensions of human behavior, such as emotion, interaction with other people, culture, and history (Bannon, 1990). Thus, the second wave of HCI recognized that user behavior “revolves
around understanding and developing task goals, responding opportunistically to intriguing system events, and planning, carrying out and evaluating courses of action” (Carroll, 2000, p. 57). From this perspective, “users” are no longer perceived as “subjects” but instead as fully functioning human beings, who need to be understood in terms of their individual skill sets, shared practices, past experiences, goals, and objectives (Bannon, 1990; Bannon, 1991; Carroll, 2000; Connolly, et al., 2008).

More recently, the third wave of HCI recognizes that use contexts typically consist of multiple users and multiple applications, which expands the notion of users to include their emotions and the quality of their interactive experiences (Bødker, 2006). As a result, “the ‘user’ is reconstructed as a creator of actions who precedes an identity rather than a user with an identity that prescribes actions” (Wakkary & Maestri, 2008, p. 13). While this view of the user is prevalent in third wave HCI, the importance of holistically understanding the user may have been first suggested by Hansen (1971). As discussed previously, Hansen’s first principle of “user engineering” was “know thy user” and he encouraged designers to understand the user’s “education, experience, interests, how much time he has, his manual dexterity, the special requirements of his problem, his reaction to the behavior of the system, his patience” (Hansen, 1971, p. 527). This research considers users in terms of three characteristics that have been shown to influence system use: users’ experience with computers, experience with the interfaces, and the quality of their domain knowledge (Nielsen, 1994; Grossman, Fitzmaurice, & Attar, 2009).
2.3.3.2 The Task

The predominant view within HCI is that a user’s context is a dynamic construct that is both determined by and a determinant of human activity. For instance, Crabtree et al. (2009) noted that “however different new technologies may be, what people do through them may not…if the technology does not support familiar activities its actual use can become problematic” (p. 886). Similarly, Norman (2005a; 2005b; 2006) advocated for Activity-Centered Design (ACD) as an alternative to Human-Centered Design (HCD) because it would force designers to understand the activities a product is designed to support rather than just the users who will be using it. A successful product, in Norman’s view, is one that fits into the requirements of the underlying activity in an easily understandable way. Norman’s motivation for developing the ACD approach came from the concept of the “taskonomy” which was first described by Dougherty and Keller (1982) in their study of blacksmiths. Dougherty and Keller found that the linguistic sorting of tools into similar groups (e.g., hammers, saws, clamps, etc.) was illogical and irrelevant to blacksmiths who instead organized their tools based on the tasks commonly associated with those tools. Along these lines, Denning and Dargan (1996) argued for Action-Centered Design by noting that “customers are more likely to be satisfied by software that is transparent in their domain of work, because it allows them to perform similar actions without distraction, and to perform new actions about which previously they could only speculate” (Denning & Dargan, 1996, p. 110). The importance of understanding users’ tasks is also underscored by several theoretical frameworks germane to HCI: situated action, activity theory, and distributed cognition.
The concept of situated action was first proposed by Suchman (1987), who argued that actions are not the result of carefully articulated plans but rather a dynamic process by which a person navigates his or her circumstances in order to achieve some specific goal. From this perspective, understanding context is a matter of understanding what a user is trying to do within a physical space. Dourish (2004) argued that the situated action perspective turns context into an “interactional” problem because “it is not simply the case that something is or is not context; rather, it may or may not be contextually relevant to some particular activity” (p. 22). The interactional view assumes that context is a dynamic, unpredictable relationship between actions and objects because context simultaneously defines and is defined by the activities taking place within it (Dourish, 2004). As such, the situated action view of context focuses on “the practical, everyday, ordinary achievements and actions of members of a particular society” (Räsänen & Nyce, 2006, p. 178). Although the situated action perspective acknowledges that artifacts and social constructs are important elements of action, the “true locus of inquiry” (Nardi, 1996b, p. 71) is the action itself.

An alternative perspective comes from activity theory. Nardi (1996a) describes activity theory as a “research framework and set of perspectives” (p. 7) originating in the Soviet Union in the 1920s as an extension of Vygotsky’s cultural-historical psychology. According to Kuutti (1995), the three key principles of activity theory are 1) activities are the basis units of analysis, 2) activities have a history and are under constant development, and 3) artifacts take a mediating role in activities. Activity theory “claims that activity defines context, where an activity comprises a subject (the person or group doing the activity), an object (the need or desire that motivates the activity), and
operations (the way an activity is carried out)” (Greenberg, 2001, p. 260). It also asserts that human activity is a three-level hierarchy consisting of activities (with associated motives), actions (with defined goals), and operations (which occur under specific conditions) (Kuutti, 1995; Bertelsen & Bødker, 2003). In this sense, designing an artifact is less about designing a “thing” and more about designing new conditions under which an activity can be performed (Bannon & Bødker, 1991). There are many other important aspects of activity theory (some will be covered in later sections), but the most salient point from this discussion is that, from the activity theory perspective, understanding human activity is critical to understanding context.

A third relevant theoretical framework for this research is distributed cognition. The theory of distributed cognition was developed primarily by Ed Hutchins based on ethnographic work studying ship navigation in the United States Navy; specifically, Hutchins observed how the operators, engineers, and technicians worked together to steer the ship and was interested in finding out what they knew and how they knew it. Like traditional theories of cognitive science, distributed cognition seeks to understand the structure of cognitive systems. However, the main difference is that distributed cognition “extends the reach of what is considered cognitive beyond the individual to encompass interactions between people and with resources and materials in the environment” (Hollan, Hutchins, & Kirsh, 2000, p. 175). This premise alters the meaning of cognition because it focuses on “the integration of cognition with action” (Hutchins, 1995, p. 368). In this manner, cognition is created by the coordination of objects (people and artifacts) within a functional system, which creates changes both inside and outside of the individuals who are involved. Unlike traditional cognitive science, the distributed
cognition perspective treats cognition as a process that is deeply embedded in an activity rather than as a process that takes place within an individual’s mind (Hutchins, 1995). For this reason, some researchers have equated the concept of a “cognitive system” with the notion of “activity” from activity theory (e.g., Nardi, 1996b). Other elements of the theory of distributed cognition will be covered in later sections, but the most relevant point from this discussion is that, from the distributed cognition perspective, cognition arises from and is an elemental part of human activity.

In addition to the theories described above, several HCI researchers have described the importance of understanding users’ tasks. For example, Carroll (2000) argued that “computers… unavoidably restructure human activity, creating new possibilities as well as new difficulties” (Carroll, 2000, p. 44). Elsewhere, Redström (2008) argued that the ideal design process should focus on what people do rather than on who they are and advocated for a shift from “thing-design” to “use-design”. This emphasis on tasks is best exemplified by Wakkary and Maestri (2008), who argued that designers must focus on what technology can do rather than who can use technology: “in considering a chair one does not model a chair user, rather a designer thinks of what actions can be performed with a chair like stacking, sitting, lying it on its side, hanging it, hanging things from it, putting things on it, tilting, and so on” (p. 13). There are different perspectives on what constitutes a “task” but this research considers “task” in a traditional HCI sense as the actions users are trying to complete with a system. In this way, tasks are broader than operations, which refer to atomic level movements or gestures, but narrower than activities, which may require completing multiple tasks or actions over a period of time (Nardi, 1996b).
2.3.3.3 The Physical Environment

The final element of context is the environment in which the user-system-task interaction takes place. This element of context can be broken down further into two components: the physical environment and the social/cultural environment.

Every user-system interaction takes place in a physical location that “includes the larger world outside of the user and the system directly being used” (Winograd, 2001, p. 405). The physical environment element of context asserts that the physical space surrounding a user-system interaction plays a defining role in that interaction. As one example, Crabtree et al. (2009) noted the importance of this physical component of context when they observed that “there is broad recognition… that the context of system design is changing. The workplace has ceased to be the exclusive focus of design. The digital, like a host of technologies before it, is permeating society at large” (p. 880). Connolly et al. (2008) also noted the importance of the physical world in defining context. They argued that the physical context was part of both the “narrower” and “broader” situational context: the “narrower” physical context was labeled the “setting” and was composed of time, place, environmental considerations (e.g., temperature, humidity, lighting, etc.), computer hardware, computer software, network connectivity and bandwidth; the “broader” situational context consisted of the universe, the world, and geographic areas of the world.

There is a clear relationship between the task and the physical environment in which the task is carried out. As Winograd (2001) noted, “context is an operational term: Something is context because of the way it is used in interpretation, not due to its inherent properties…Features of the world become context through their use” (p. 405).
So although Norman (2006) argued that design should be centered on activities rather than users, he also recognized the importance of the physical environment in determining what those activities may be: “once an activity has begun, then taskonomy is the way to go, where things used together are placed near one another, where any one item might be located logically within the taxonomic structure but also wherever behaviorally appropriate for the activities being supported” (p. 63). This point was also made by Dougherty and Keller (1982) in their description of a taskonomy for blacksmiths. They found that in addition to being categorized by task, blacksmiths also conceptualized tools based on where the tool was located in the shop (e.g., near the stump, near the fire, etc.). From this perspective, the physical environment plays a pivotal role in shaping the actions that take place within it.

The importance of the physical environment is also one of the basic tenets of the theory of distributed cognition. As noted previously, the primary assertion of this theory is that cognition arises from human activity, where activity is considered the coordination between people and artifacts. Hutchins (1995) noted that while navigating the ship, the ship's quartermasters manipulated numbers and symbols, drew lines on a chart, and created a number of other physical artifacts. As a result, the very act of navigating the ship created changes in the “material environment that the quartermasters share with and produce for each other” (p. 360). In another example, Hollan et al. (2001) noted that materials people use in their work are not just seen as stimuli but rather as “elements of the cognitive system itself. Just as a blind person’s cane or a cell biologist’s microscope is a central part of the way they perceive the world, so well-designed work materials become integrated into the way people think, see, and control activities, part of the
distributed system of cognitive control” (p. 178). As a result, the theory of distributed cognition recognizes that a cognitive system is often the result of coordination between internal objects (within the mind) and artifacts located in the surrounding physical environment.

To a lesser extent, the physical world is central to the theory of situated action as well. As Nardi (1996b) noted, the unit of analysis in situated action is neither an individual nor the environment but rather “a relation between the two” (p. 71). From this viewpoint, actions take place in “a complex world of objects, artifacts, and other actors, located in space and time” and these factors are “seen as the essential resource that makes knowledge possible and gives action its sense” (Suchman, 1987, p. 179). As a result, situated actions must be interpreted in terms of the “setting” and the “arena”, where the setting is the relationship between the actors and the physical environment and the arena is “a stable institutional framework” (Nardi, 1996b, p. 71). Take, for instance, Suchman’s (1987) famous example of the Trukese navigator. Unlike the European navigator who creates and executes a purposeful plan, the Trukese navigator identifies an objective (e.g., travelling to a specific location) and adjusts his course along the way based on the material circumstances of his environment (e.g., waves, rocks, winds, etc.). In this way, the “organization of situated action is an emergent property of moment-by-moment interactions between actors, and between actors and the environments of their action” (Suchman, 1987, p. 179).

The physical element of context is also one of the most commonly cited elements of context in the field of Context-Aware Computing, where the predominant view is that context is primarily a function of place (Dey & Abowd, 1999). Dourish (2004) pointed
out that context from this perspective “describes features of the environment within
which the activity takes place, but which are separate from the activity itself” (Dourish,
2004, p. 22). In one of the first descriptions of Context-Aware Computing, Schilit and
Theimer (1994) discussed the notion of “active maps” as a way of tracking the location of
users and other objects within an environment. Similarly, Voida et al. (2002)
differentiated “physical context” from “virtual context”, where physical context was
measured by tracking users’ location relative to the public areas of the office or near
common devices such as the printer. However, Dey et al. (2001) expanded the notion of
location to include elevation, orientation, and co-location or proximity of entities. Other
researchers have also included other elements of location, including the time of day, day
of the week, season, temperature, ambient light, noise level, etc. (Dey & Abowd, 1999;
Dey, 2001; Dey et al., 2002). The general consensus is that the location of users and
objects is paramount to building context-aware applications, implying that “setting-
aware” (Winograd, 2001, p. 406) might be a more appropriate term to describe these
types of systems. For simplicity, this research will only consider the physical
environment in terms of location and time.

2.3.3.4 The Social/Cultural Environment

Finally, in addition to the physical environment in which a user interacts with a
system, researchers have also noted that there are important social and cultural factors
that also influence a user-system interaction (Bødker, 2006). This observation is
fundamental to the third wave of HCI because it focuses on meaning-making rather than
problem-solving, which “promotes a view towards situated and emergent properties of
interaction” (Ylirisku, et al., 2009, p. 1131). This view is sometimes referred to as
“embodied action” because, unlike situated action, it associates action with meaning and asserts that “elements like structure, history, and culture have [a role] in rendering a social world meaningful to those who inhabit it” (Räsänen & Nyce, 2006, p. 178).

The social/cultural environment is also included as a key element of activity theory. Although the activity theory perspective is centered on human activity, it recognizes that activities are embedded within a socially structured environment (see Figure 3). Specifically, activity theorists maintain that “consciousness is located in everyday practice: you are what you do. And what you do is firmly and inextricably embedded in the social matrix of which every person is an organic part” (Nardi, 1995a, p. 7). In this “social matrix”, the subject-object relationship is mediated by the community (or the people who share the same object), the relationship between the subject and the community is mediated by rules (norms, conventions, social relations, etc.), and the relationship between the community and the object is mediated by the division of labor (the organization of a community, which may be implicit or explicit). This complex web of relationships is typically depicted in the form of a triangle, as shown in Figure 3 (Kuutti 1995; Bertelsen & Bødker, 2003). Because an activity system consists of objects, artifacts and people, “context is not just 'out there.' Context is both internal to people…and, at the same time, external to people, involving artifacts, other people, specific settings” (Nardi, 1995b, p. 76).
The theory of distributed cognition also affirms that the social/cultural environment plays a major role in defining a cognitive system. The distributed cognition perspective recognizes that cognition is shared within and among social groups, but it also “includes phenomena that emerge in social interactions as well as interactions between people and structure in their environments” (Hollan, et al., 2000, p. 177). Distributed cognition is as much about social interactions as it is about human activity, and it asserts that cognition is inseparable from culture. As Hutchins noted, distributed cognition is an “integrated view of human cognition in which a major component of culture is a cognitive process…and cognition is a cultural process” (Hutchins, 1995, p. 354). As a result, culture and cognition have a symbiotic relationship in which culture both emerges from the process of human activity (and, thus, cognition) while simultaneously shaping those cognitive processes (Hollan, et al., 2000). For this reason, Hutchins (1995) noted that traditional theories of cognitive science view cognition as a “physical-symbol-system” in which “the hands, the eyes, the ears, the nose, the mouth,
and the emotions” (p. 363) of the individual have been stripped away; instead, he argued that cognition must be viewed as a “sociocultural cognitive system” where “culture, context, and history...are fundamental aspects of human cognition” (Hutchins, 1995, p. 354).

Similarly, in addition to the physical component of situational context, Connolly et al. (2008) identified a socio-cultural component: the “narrower” socio-cultural context was labeled the “scene” and consisted of the participants and their social attributes (e.g., their roles, attitudes, and level of computer literacy), the social circumstances (e.g., informal conversation, formal conference, etc.), and both the intended and actual outcomes of the discourse; the “broader” socio-cultural context was composed of the social structure (groups, categories, organizations), cultural norms (beliefs, values, ideologies, laws, conventions, etc.), economics (cost), history, and the available modes of communication (Connolly, et al., 2008). Again, while any or all of these aspects may comprise the socio-cultural aspect of context, it is up to the designer to determine which aspects are relevant in a given situation. Recognizing the importance of the social/cultural environment can help designers go beyond moment-by-moment actions of individual actors and help to interpret, rather than discover, the structures that underlie and provide meaning for human activity (Räsänen & Nyce, 2006; Ylirisku, et al., 2009). While there are many aspects of the social environment which may influence system use, this research considers only those aspects which can be simulated in a laboratory setting: the motivation for completing a task, including the goal that must be achieved, and the social incentives for completing of a task, including educational, professional, or personal expectations.
2.3.4 Summary

Three studies were found that explicitly made the connection between usefulness and context but they each used different definitions of usefulness and context and each employed different experimental methods, which limited their overall impact and applicability to the study of usefulness. Although context has been defined and conceptualized in a variety of ways, this literature review identified four major contextual elements: the user, the task, the physical environment and the social/cultural environment.

2.4 Usefulness and Usability

Researchers in HCI have long recognized the close connection between usefulness and usability (e.g., Grudin, 1992) but there have been few attempts to empirically study this relationship. The following section will explore the complex relationship between usefulness and usability. First, usability will be discussed in terms of how it is defined and measured. Next, the concepts of perceived ease of use and perceived usefulness will be investigated. Finally, the relationship between usability and usefulness will be examined.

2.4.1 Defining and Measuring Usability

One of the biggest obstacles to usability research is deciding on a standard definition of the term. As noted by Gray & Salzman (1998) in their review of UEM studies, “the most important issue facing usability researchers and practitioners alike...[is] the construct of usability itself” (p. 238). In their seminal paper, Gould and Lewis (1985) discussed four principles of “designing for usability” but did not provide any definition of the term. Although usability research dominated HCI literature throughout
the ensuing decade, it was the publication of the ISO standard in 1997 that finally
“elevated the practice of usability to a level of acceptability that had previously been
difficult to achieve” (Lindgaard & Parush, 2008, p. 223).

This observation was confirmed by Chen et al. (2009) who found that there were
no formal definitions of usability published prior to 1989. Overall, Chen et al. found 63
different definitions of usability containing 502 different terms or concepts. Using
standard content analysis procedures, the concepts were broken down further into 11
categories: learnability, efficiency, effectiveness, memorability, error tolerance,
attitude/satisfaction, usefulness, control/flexibility, user characteristics, context/purpose,
and interface/design. The five most frequently cited concepts – accounting for 77% of the
total number of concepts – were learnability, effectiveness, user characteristics,
attitude/satisfaction, and efficiency (Chen et al., 2009). From another perspective,
McGee, Rich, and Dumas (2004) proposed a “user-defined definition” of usability as “[a
user’s] perception of how consistent, efficient, productive, organized, easy to use,
intuitive and straightforward it is to accomplish tasks within a system” (p. 909).
However, the two most frequently cited definitions of usability are Nielsen (1993) and
the ISO (1997), and other widely cited definitions are Shackel (1991) and Shneiderman
(1998). These four definitions are presented in Table 2 (a similar table is presented by
van Welie, van der Veer, & Eliëns, 1999).
Table 2. Definitions of usability.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexibility</td>
<td>Efficiency</td>
<td>Speed of performance</td>
<td>Efficiency</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>Errors</td>
<td>Rate of errors by users</td>
<td>Effectiveness</td>
</tr>
<tr>
<td>Learnability</td>
<td>Learnability</td>
<td>Time to Learn</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Memorability</td>
<td>Retention over time</td>
<td></td>
</tr>
<tr>
<td>Attitude</td>
<td>Satisfaction</td>
<td>Subjective Satisfaction</td>
<td>Satisfaction</td>
</tr>
</tbody>
</table>

While there seems to be little consensus about a precise definition of usability, there seems to be agreement that usability refers to the effectiveness, efficiency and satisfaction with which users can complete tasks. Given the diversity of perspectives and purposes of usability evaluation research, use of these measures is surprisingly consistent: in a review of 180 usability studies, Hornbæk (2006) found that 78% of the studies either contained a measure of effectiveness or controlled for effectiveness, 81% reported a measure of efficiency or controlled for efficiency, and 62% reported a measure of satisfaction; however, only 13% of the studies reported all three measures. Similarly, there are five commonly used metrics associated with the measurement of usability⁴: time to complete tasks, error rate, accuracy, task completion rate, and satisfaction (Sauro & Kindlund, 2005; Hornbæk & Law, 2007; and Sauro & Lewis, 2009). In a separate study of 90 distinct usability tests carried out between 1983 and 2008, Sauro and Lewis (2009)

---

⁴ These metrics are nearly identical to the metrics used in the evaluation studies from the 1960s and 1970s. While the persistence of these metrics may be evidence of their evaluative power, it is still striking that, in over 40 years of evaluation research in HCI, the metrics used to evaluate systems have remained virtually unchanged.
found that 99% of the usability tests measured task completion time, 98% measured completion rate, 58% measured error rate, and 48% measured post-test satisfaction.

This focus on speed, effectiveness and satisfaction is largely driven by the widespread adoption of the ISO definition of 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” (ISO, 1997, Sec. 3; italics added). This definition of usability is remarkably similar to the definition of usefulness proposed in this paper as “the extent to which a system’s functions allow users to complete a set of tasks and achieve specific goals in a particular context of use.” However, the key difference is that usefulness simply refers to the extent to which a system fits a specific context while usability states that it does so “with effectiveness, efficiency, and satisfaction.” This is a critical distinction between the two concepts, although it is clear that usability and usefulness are still closely related.

2.4.2 Usability, Perceived Ease of Use and Perceived Usefulness

Over the last two decades, the constructs of perceived usefulness and perceived ease of use have been widely used to study acceptance of information systems as part of the Technology Acceptance Model (TAM). Researchers have used the TAM to evaluate over 50 different types of systems, including e-mail, groupware, expert systems, voice mail, calculators, digital libraries, spreadsheets, and electronic health systems. The construct of perceived usefulness is consistently found to be highly correlated with actual systems use (Venkatesh, Davis, & Morris, 2007). First described in Davis (1989) and Davis, Bagozzi, and Warsaw (1989), the TAM is “considered the most influential and commonly employed theory for describing an individual’s acceptance of information
systems” (Lee, Kozar, and Larsen, 2003, p. 753). In their meta-analysis of TAM studies, Lee et al. (2003) found that the first two articles describing the TAM received almost 700 citations between 1989 and 2003 and subsequent research has noted that citations of the two TAM articles are now well over 1000 (Venkatesh, et al., 2007).

Presented as a special case of the Theory of Reasoned Action (TRA), the TAM theorizes that actual computer usage is determined by behavioral intention, which in turn is “jointly determined by the person's attitude toward using the system (A) and perceived usefulness (U)” (Davis et al., 1989, p. 985). The major assumption of the TAM is that attitude toward using a computer system and behavioral intention to use a computer system are both primarily determined by the construct of perceived usefulness which is mediated by the construct of perceived ease of use, as indicated in Figure 4.

![Figure 4. The Technology Acceptance Model (Davis, Bagozzi, & Warsaw, 1989)](image)

These constructs are measured by administering a questionnaire to users after a brief interaction with a system. For the purposes of the TAM, Davis defined perceived usefulness (PU) as “the degree to which a person believes that using a particular system would enhance his or her job performance” and perceived ease of use (PEOU) as “the
degree to which a person believes that using a particular system would be free of effort” (Davis, 1989, p. 320). The original validated six-item scale for measuring perceived usefulness is presented in Table 3.

<table>
<thead>
<tr>
<th>Table 3. Measurement scales for perceived usefulness (Davis, 1989).</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Using [the system] in my job would enable me to accomplish tasks more quickly.</td>
</tr>
<tr>
<td>3. Using [the system] in my job would increase my productivity.</td>
</tr>
<tr>
<td>4. Using [the system] would enhance my effectiveness on the job.</td>
</tr>
<tr>
<td>5. Using [the system] would make it easier to do my job.</td>
</tr>
<tr>
<td>6. I would find [the system] useful in my job.</td>
</tr>
</tbody>
</table>

The construct of Perceived Usefulness (PU) presented in the TAM captures some of the components—namely, efficiency and effectiveness—commonly found in the traditional HCI definitions of usability, as described in the previous section. Further, the fundamental assumption of TAM is that a system’s usefulness is entirely determined by the system’s effect on user performance, particularly in a workplace or job-related setting. In later studies using the TAM, Davis (1993) explicitly stated that “perceived usefulness concerns the expected overall impact of system use on job performance” (p. 477). This emphasis on performance is intentional, as Davis (1989) noted that the theoretical foundations of the TAM are rooted in the notion that “performance” is most often correlated highly with actual systems use. Davis’ research into other areas (cost-benefit analysis, adoption of innovations, and evaluation of information reports, among others) also provided strong indications that PU is closely tied to performance (Davis, 1989). Researchers in the IS field have noted that TAM’s emphasis on performance may be a limitation of the model, and have proposed no less than 25 additional variables (such
as social presence, job relevance, accessibility, prior experience, etc.) which also explain some of the variance in users’ perceptions of usefulness, perceptions of ease of use, and actual usage behavior (Lee et al., 2003). By focusing on performance-related issues in terms of effectiveness and efficiency, Perceived Usefulness (PU) as defined in the TAM is quite similar to the HCI construct of usability and thus distinct from the concept of usefulness proposed by this research.

2.4.3 The Relationship between Usability and Usefulness

Although usability and usefulness refer to two clearly distinct aspects of systems use, the relationship between the two is not easily apparent. Some usability researchers identify usefulness as one aspect of usability. Gulliksen et al. (2003) argued that the ISO definition of usability “includes the concept of utility or usefulness, often seen as separate from usability” [italics in original] (p. 407). Frandsen-Thorlacius et al. (2009) observed that “recent definitions [of usability] have extended the concept to also include aspects of utility, experience, fun, and culture” (p. 42). Similarly, in the QUIM (Quality in Use Integrated Measurement) model proposed by Seffah et al. (2006), usefulness was listed as one of the ten components of usability. In their meta-review of usability definitions, Chen et al. (2009) also identified usefulness as one of ten commonly cited attributes of usability although they found it was actually the least commonly cited attribute, appearing in just 6 (9.5%) of the 63 usability definitions they reviewed. The latter finding seems indicative of a broader trend within HCI to view usability and usefulness as two separate (but closely related) constructs.

Researchers in HCI have long recognized that usefulness is more than just usability. This perspective is reflected in English et al.’s (1967) aforementioned landmark
study of input devices, where they noted that there are many other elements of the interaction which need to be taken into consideration beyond simply ease of use. They noted “the details of the usage system in which the device is to be embedded make too much difference… [and] the tradeoffs for the characteristics of fatigue, quick transfer to and from a keyboard, etc., will heavily weight the choice among the devices” (English et al., 1967, p. 14-15).

From a broader perspective, the distinction between usefulness and usability can be demonstrated by looking at the number of systems that are highly un-useable, but are successful because they are useful. As noted by Greenberg and Buxton (2008), the World Wide Web is a perfect example because although it suffered from numerous usability problems at the start, it was still highly used and became much more usable over time. Conversely, there are many systems that are unused despite having high usability. The problem in most of these cases is that the systems have low usefulness. In one study of how users adopted an electronic mail program called PMail, Maryniak-Nelson and Caldwell (1992) found that although users were satisfied with the usability of PMail, they continuously judged the appropriateness of using the technology based on organizational and situational factors (e.g., whether the intended recipient of the message had access, whether a reply was urgent, etc.). They concluded that “while the software was ‘easy to use’ it was not used very much because the functionality it provided did not match the needs of the organization” (Maryniak-Nelson & Caldwell, 1992, p. 879). In one attempt to describe this phenomenon, Keil, Beranek, and Konsynski (1995) proposed the Usefulness/Ease of Use Grid (Figure 5).
According to Keil et al., the goal of any designer is to create a system with both high usefulness and high ease of use. They classified these systems as SuperTools (Quadrant IV) while systems with low usefulness and low ease of use were termed Rejects (Quadrant I). They adopted this framework as part of a case study of the CONFIG system. The developers of the CONFIG system noticed that users were not using the system even though they had recently overhauled the interface to make it more user-friendly. Keil et al. found that the primary problem with CONFIG was that it had poor task/tool fit. Designers of the tool misunderstood how the intended users were going to use the system, which resulted in a system that was poorly matched to users’ needs. This finding led Keil et al. (1995) to conclude that the CONFIG system exhibited Low Usefulness (Quadrant I) to begin with, and that “no amount of EOU [ease of use] will compensate for low usefulness” (p. 89).

In short, useful products are adopted and used by people, often regardless of usability problems. For example, in their aforementioned study investigating the
usefulness of an enterprise resources planning (ERP) system, Abdelnour-Nocera et al. (2007) found that users defined the system’s usefulness in terms of their coping strategies for dealing with the usability problems they experienced. While some users invented workarounds to circumvent the system in cases where they deemed it unsuitable for the task, others simply changed their work practice to conform to the system’s underlying business logic. Users typically viewed the workarounds as temporary solutions, and they used these experiences to inform developers of how the system could be improved. In other words, the system’s usefulness was the primary driver for improving its usability. This finding suggests that usability often comes after usefulness, not before. “A novel and useful innovation (even though it may be hard to use) is taken up by people, and then competition over time forces that innovation to evolve into something that is more usable” (Greenberg & Buxton, 2008, p. 116).

An alternative viewpoint was posed by Petroski (1994), who argued that useful tools often emerge when people recognize and fix usability problems:

“What form does follow is the real and perceived failure of things as they are used to do what they are supposed to do. Clever people in the past, whom today we might call inventors, designers or engineers, observed the failure of existing things to function as well as might be imagined. By focusing on the shortcomings of things, innovators altered those items to remove the imperfections, thus producing new, improved objects.” (Petroski, 1994, p. 20)

There are other cases when the usability problems are so severe that they may cause users to give up trying to use it, jeopardizing users’ ability to recognize the usefulness of the system (Long, et al., 1983). This observation likely led Dillon & Morris (1999) to define usability as “the extent to which users can exploit the utility of the system” (p. 6) and Goodwin (1987) to remark that “at best, a system with poor usability
will cost its users time and effort; at worst, it will not be used at all, and its functions may be removed because their utility has not been demonstrated” (p. 232). This viewpoint was also expressed by Eason (1984), who noted the role of usability as a mediator of system utility. In Eason’s view, users will only see the benefit of using all of the functions offered by a system (i.e., the system’s usefulness) if they perceive those functions as being easy to use (i.e., the system’s usability). In this way, “usability serves to limit the degree to which potential utility becomes actual utility” (Eason, 1984, p. 133). Eason developed this framework as part of a study attempting to determine why users were not taking advantage of all the functions provided by an on-line inquiry system for bankers. In preliminary tests, he observed that even though the system provided 36 codes corresponding to common banking tasks, 75% of the functions used by the bankers consisted of just four codes. In a follow-up study, he found that bankers could recognize the “correct” codes just 53% of the time. Eason concluded that there were significant usability barriers to learning and using all 36 codes, which ultimately had a negative impact on the overall usefulness of the system. However, Eason also found that 26% of the codes used by the bankers were technically “incorrect” but still likely resulted in a successful outcome. Despite these apparent usability problems, the bankers were able to successfully complete the task almost 80% of the time, which suggests that the usability problems he identified may not have had a huge impact on system usefulness.

2.4.4 Summary

A key difference between usability and usefulness is that usefulness refers to how well a system helps users complete their tasks within a specific context of use while usability refers to whether it does so with “effectiveness, efficiency, and satisfaction.”
The Technology Acceptance Model (TAM) presents one way of looking at the difference between the two concepts, but TAM’s Perceived Usefulness (PU) construct focuses on work-related performance in terms of effectiveness and efficiency, which closely mirrors the concept of usability as commonly viewed in HCI. There seems to be agreement within the field of HCI that usefulness cannot be assumed to be an attribute or component of usability, and many researchers have theorized about the difference between the two concepts. However, there have been few efforts to empirically determine how and to what extent usability and usefulness are related. Therefore, this research will attempt to provide empirical data regarding the exact nature of this relationship.

2.5 Usefulness, Aesthetics and Enjoyment

In addition to usability, there is a strong possibility that usefulness is influenced by subjective factors related to the visual design of the system and the emotions it arouses. This section will discuss the concepts of aesthetics and enjoyment and their possible relationship to usefulness.

2.5.1 Aesthetics

The debate between form and function has been prevalent in industrial design for over sixty years. For example, although Dreyfuss (1955) describes the artist as an “unfortunate creature” who “disfigured so much of our industrial product,” he also noted that the modern industrial designer should strive for “handsome surface appearance” (p. 78). But up until the mid-1990s, there was a common perception that “mainstream HCI... either belittles the importance of aesthetics or ignores it altogether” (Tractinsky, 1997, p. 116). In an attempt to reverse this trend, Kurosu and Kashimura (1995a; 1995b) reported the results of two experiments investigating the relationship between aesthetics and
usability. In the first experiment, participants rated the aesthetic appeal and the apparent usability of 26 different ATM layouts. In the second experiment, participants were asked to complete a series of tasks using 8 representative ATM layouts from the previous study and rate the aesthetic appeal of each interface before and after completing the tasks. In both studies, the researchers found evidence of a concrete relationship between aesthetics and usability. More importantly, the results suggested that users “may be strongly affected by the aesthetic aspect of the interface even when they try to evaluate the interface in its functional aspects” (Kurosu and Kashimura, 1995a, p. 293). In effort to replicate these results, Tractinsky (1997) designed three experiments using the exact same ATM layouts and measurement instruments Kurosu and Kashimura employed. Although Tractinsky hypothesized that culture was a predominant factor in Kurosu and Kashimura’s findings, he found similar results in all three studies: users’ ratings of aesthetics were highly correlated with both perceived and experienced usability (Tractinsky, 1997).

Despite the efforts described in these early studies, the concept of aesthetics received little attention in the HCI literature until Tractinsky et al. (2000) observed that not only were users able to distinguish aesthetics from other system attributes, but their perceptions of aesthetics were (again) highly correlated with perceptions of usability. Tractinsky et al. also showed that users’ perceptions of aesthetics and usability were highly correlated with overall satisfaction, leading them to conclude that “what is beautiful is usable.” Similarly, De Angeli, Sutcliffe, and Hartmann (2006) found a strong correlation between evaluations of usability and aesthetics and discovered a type of “halo effect” in which users seemed to disregard usability problems if the aesthetics were rated
highly. This result echoed another finding by Tractinsky et al. (2000), which was that
users' high perceptions of aesthetics decreased after interacting with the system while
users' low perceptions of aesthetics increased after interacting with the system.

Other studies have confirmed the existence of this relationship. For example,
Hartmann et al. (2007) found that the task and users’ background also influenced
perceptions of aesthetics and that “good aesthetic design can overcome some deficit of
usability problems” (p. 395). van Shaik and Ling (2009) found that context and usage
increased the stability of aesthetic judgments over time, but their study found no
significant correlations between usability and aesthetic value, contradicting previous
findings. De Angeli, Hartmann, and Sutcliffe (2009) found that brand attitude was the
most important predictor of behavioral intentions to revisit or recommend the website to
a friend, while expressive aesthetics, usability, and classical aesthetics were also strong
predictors. In summary, studies have consistently shown that the aesthetic appeal of a
system has some type of mediating effect on the construct of usability, leading to the
strong probability that usefulness may be similarly influenced.

2.5.2 Enjoyment

As noted by Carroll and Thomas (1988), the traditional HCI focus on usability or
ease of use likely comes at the expense of subjective factors like fun or enjoyment. They
argued that (at the time) HCI provided “no understanding of how people feel about or
relate to tools they use. We do not even have a framework for posing such questions” (p.
23). As HCI embraces the concept of UX and enters its third wave, where the line
between work and personal computing is blurred (Bødker, 2006), there is an increased
need to concentrate on emotions and experience because user satisfaction is becoming an
increasingly complex construct composed of several affective factors (Lindgaard & Dudek, 2003; Bødker, 2006).

To address this challenge, evaluation studies have been expanded to include these emotional and affective factors. As an early example, Malone (1982) proposed heuristics for designing enjoyable user interfaces. Culled from his experience with video games, Malone proposed that an enjoyable interface has three main qualities: challenge, fantasy, and curiosity. First, an interface can challenge users by providing a clear goal, offering performance feedback, and providing uncertain outcomes (e.g., variable difficulty levels, layers of complexity, and multiple level goals). Second, an interface can appeal to users by providing an emotionally appealing fantasy. Finally, an enjoyable interface must appeal to users’ innate sense of curiosity by providing an optimal level of information complexity, taking advantage of audio/visual effects, randomness, and humor, and by taking advantage of users’ desire to fulfill their own knowledge gaps, needs, or inconsistencies (Malone, 1982).

More recently, Hart et al. (2008) found that although Facebook performed poorly in a heuristic usability evaluation, a majority of users reported the site as being “easy” or “very easy” to use. To explain this phenomenon, the researchers noted that participants most often cited affective factors (e.g., “curiosity” or “enjoyment”) to describe their positive experiences interacting with the site. Lindgaard and Dudek (2003) found that a website with low usability and high aesthetic appeal could provide a similarly pleasant experience to a website with high usability and high aesthetic appeal, even if those experiences were qualitatively different. Other examples of this type of research include De Angeli et al. (2009), who reported that users with negative brand attitudes also
perceived websites as less attractive and the experience as less pleasurable than users with neutral or positive attitudes of the brand. Joinson (2008) uncovered seven reasons why people use the popular social networking site Facebook, all of which focused on experiential (rather than usability) outcomes: social connections, shared identities, posting and viewing photographs, content within Facebook (applications), social investigation, social network surfing, and status updates (note that none of these reasons address issues related to usability). In each of these cases (and many others), researchers continue to find strong evidence that subjective factors have a strong influence on usability, which suggests that the usefulness may also be influenced by subjective factors related to emotion and enjoyment.

2.5.3 Summary

In addition to the practical and pragmatic components of system use, there are also important emotional and affective factors that influence user-system interactions. Many researchers have found the existence of a strong correlation between the aesthetics of a system and its usability, suggesting the possibility of a similar relationship with usefulness. There is also a broad range of research that seeks to understand how users’ perceptions are influenced by their emotional and affective states, including feelings of enjoyment and fun. However, there is no knowledge of how affect and emotion may influence perceptions of usefulness. This research will attempt to fill this gap.

2.6 Summary and Research Questions

Usefulness has implicitly played a role in evaluation research over the last several decades, but the meaning of the term “usefulness” has changed over time, from a system-based focus on reliability to a user-centered focus on learnability and ease of use. Today,
user experience (UX), which includes both practical and emotional aspects of using systems, is beginning to replace usability as the dominant research paradigm in HCI, implying that the meaning of usefulness may need be revised accordingly. Consequently, this research is guided by the following question: *what does usefulness mean for modern interactive computer systems?*

Designing useful systems has long been cited as a goal of employing user-centered design methods, but the study of usefulness as a construct has been largely ignored in HCI and the term *usefulness* has been defined and used in many different ways. However, most researchers have used the term to describe one or more of the following: the functions provided by the system, the tasks users are trying to complete, the goals users are trying to achieve, and the context in which the system is being used. A working definition of usefulness is proposed as “the extent to which a system’s functions allow users to complete a set of tasks and achieve specific goals in a particular context of use.” It is widely believed that the usefulness of a system is tied to the context in which it is used, but there have been few studies that have both explicitly recognized and attempted to investigate this connection. Unfortunately, little can be concluded from these studies because they each conceptualized usefulness and context differently. This research attempts to remedy this problem by 1) providing a working definition of usefulness and 2) using a common framework for understanding context as the user, the task, the tool, and the environment (both physical and social/cultural). In addition, many researchers have discussed the relationship between usefulness and usability but few studies have examined this relationship empirically and it is unclear how the two concepts are related. There are also subjective aspects of UX related to aesthetics and
enjoyment that have been shown to influence users’ perceptions of usability, but it is unknown whether they also influence perceptions of usefulness. Consequently, four research questions are proposed. The first two questions are meant to address the gaps in the literature described in the previous sections while the last two questions are meant to broaden and enrich our understanding of usefulness. The four questions are:

**RQ1:** How and to what extent are perceptions of usefulness shaped by the context—the user, task, and environment—in which the system is used?

**RQ2:** How and to what extent do perceptions of usability, aesthetics, and enjoyment influence perceptions of usefulness?

**RQ3:** To what extent do perceptions of usefulness affect users’ overall evaluation of a system?

**RQ4:** What criteria do users identify when defining or describing a useful system?

### 2.7 Significance

Designing *useful* systems has always been at the core of HCI, even though few researchers agree on what the term “useful” means. Instead, because of its initial focus on making software systems more amenable to non-experts, system evaluation in HCI has focused primarily on system usability. It is undeniable that effective systems must be easy to use, and HCI research has made significant contributions to the design and evaluation of more usable systems. But, as Greenberg and Buxton (2008) pointed out: “If all we do is usability evaluation…we will have a modest impact by making existing things better. What we will not do is have major impact by creating new innovations” (p. 116). They listed several examples of systems – Marconi’s wireless radio, the
automobile, Sketchpad – that would have failed a standard usability evaluation despite being revolutionary innovations. However, if the field of HCI needs to focus on evaluating system usefulness, it requires as a first step a thorough understanding of what it means for a system to be useful. There is a particular need for generalizable knowledge about the nature of usefulness and its relationship to common UX attributes (usability, aesthetics, and enjoyment). It is also critical to understand how usefulness is related to the context in which a system is used so that both researchers and practitioners can develop more agile, flexible and effective UX evaluation methods. As use contexts continue to increase in complexity and users continue to become increasingly advanced in their knowledge and understanding of technology, understanding usefulness in HCI becomes a critical step in developing novel, innovative systems.
3. RESEARCH METHODS

This chapter describes the research methods used in this study. First, the previous chapter is summarized and the research questions are presented. Next, the methods are described in terms of the independent and dependent variables, the experimental design, the experimental procedure, the data collection instruments, and the data analysis methods.

3.1 Introduction

As described in previous chapters, the study of usefulness has been largely ignored in HCI even though the concept has implicitly played a pivotal role in system evaluation research over the last several decades. The term usefulness has been defined and used in many different ways, but a review of the literature revealed broad consensus that the usefulness of a system is inextricably tied to the context in which it is used. It is also likely that usefulness is closely related to (but distinct from) usability and that it may also be influenced by subjective factors related to aesthetics and enjoyment. This section describes the research methods employed by the current study, which was designed to provide empirical evidence about the relation between usefulness and context, usability, aesthetics, and enjoyment. Accordingly, the four research questions for the current study are:

**RQ1:** How and to what extent are perceptions of usefulness shaped by the context—the user, task, and environment—in which the system is used?

**RQ2:** How and to what extent do perceptions of usability, aesthetics, and enjoyment influence perceptions of usefulness?
RQ3: To what extent do perceptions of usefulness affect users’ overall evaluation of a system?

RQ4: What criteria do users identify when defining or describing a useful system?

3.2 Variables

The current study used two independent variables (CONTEXT and SYSTEM) and five dependent variables (USEFULNESS, USABILITY, AESTHETICS, ENJOYMENT, and GOODNESS). Each of these variables will be described in this section.

3.2.1 Independent Variables

3.2.1.1 SYSTEM

The first independent variable used in the study was SYSTEM, with three levels. The three levels of the SYSTEM variable were represented by three different information portal websites. Information portals were selected because 1) they provided a range of potential tasks and goals that could be exploited for an experiment, 2) they are similar enough to web search engines that users could learn to use them with minimal effort and 3) they are complex enough that each information portal is sufficiently different from its competitors. Three specific information portals were used in the current study: ipl2\(^\text{5}\) (formerly the Internet Public Library), RefSeek\(^\text{6}\), and Awesome Library\(^\text{7}\). These three portals were chosen because they had similar purposes and sufficient domain overlap but were different in many other aspects (e.g., appearance, browsing structure, search capability, etc.). Because users often draw from past experiences with similar websites

\(^{5}\) http://www.ipl.org
\(^{6}\) http://www.refseek.com
\(^{7}\) http://www.awesomelibrary.org
when evaluating new websites, RefSeek's resemblance to Google likely biased users in its favor. However, this was not seen as a problem because it more closely reflects real world interactive experiences. The experiment was carried out on live versions of the three websites over a one month period (July 2011). Screenshots of the three websites are presented in Figure 6.
Figure 6. Screenshots of ipl2, RefSeek, and Awesome Library.
3.2.1.2 CONTEXT

The second independent variable used in the study was CONTEXT, with three levels. The three levels of the CONTEXT variable were represented by three contextual scenarios: one scenario was a “no scenario” and was used as a control while the other two scenarios described two hypothetical contextual situations that differed along the major dimensions of context identified in Section 2.3: the users’ task(s) and their physical and social/cultural environments. The scenarios used in this study are presented in Table 4.

<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate</td>
<td>It's 10AM on a Thursday morning. You're sitting in the iCommons computer lab and you need to start preparing a 10-minute PowerPoint Presentation for your COM 230 class ( Techniques of Speaking) on the scientific and political issues surrounding climate change. Specifically, you need to start compiling a list of reputable sources where you can find information (facts, pictures, charts/graphs, etc.) to include on your PowerPoint slides. You need to give your presentation at the beginning of class tomorrow.</td>
</tr>
<tr>
<td>First</td>
<td>It's 8:30PM on a Monday night. You’re hanging out in your dorm room when your friend walks in and asks if you know anything about the invention of the first computer. He claims it was named “IPECAC” and was invented in 1928 at Drexel University, but you don’t believe him. After a few minutes of arguing, he makes you a bet: he will give you $20 if, in the next five minutes, you can find credible sources that prove whether the first computer really was named “IPECAC” and was invented in 1928 at Drexel University.</td>
</tr>
<tr>
<td>Control</td>
<td>No Scenario (Control)</td>
</tr>
</tbody>
</table>

To control for differences across the levels of the CONTEXT variable, the two scenarios were similar in length and complexity and were written according to the guidelines provided by Carroll (2000), who stated that scenarios should contain the following components: a setting (physical location, hardware/software configuration, and
social/organizational setting), agents or actors, who each have goals or objectives (changes the agent wishes to achieve in the circumstances of the setting), and sequences of actions and events (things actors do or that happen to them that change the circumstances of the setting). As stated above, the two scenarios differed along several dimensions to ensure that the two levels of the CONTEXT variable were likely to cause differences in participants’ behavior: the task, the physical environment, and the social/cultural environment (see Table 5).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Climate Change</th>
<th>First Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task</strong></td>
<td>Information exploration (action mode)</td>
<td>Information retrieval (goal mode)</td>
</tr>
<tr>
<td><strong>Physical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Environment</strong></td>
<td>Thursday morning in a computer lab.</td>
<td>Monday night in a dormitory room.</td>
</tr>
<tr>
<td><strong>Social/Cultural</strong></td>
<td>Find resources to include in a PowerPoint presentation for a school assignment.</td>
<td>Find information to win a $20 bet with a friend.</td>
</tr>
</tbody>
</table>

In terms of task, the scenarios reflected the difference between “action mode” and “goal mode” as described by Hassenzahl (2003). In “action mode,” goals are created dynamically and are determined by the action, making the system an “end in itself.” In “goal mode,” the goal determines the user's actions, making the system “a means to an end” (Hassenzahl, 2003, p. 39). In this study, as in the study conducted by van Shaik & Ling (2011), “action mode” corresponded to an imprecise information exploration task (Climate Change) while “goal mode” corresponded to a specific information retrieval task (First Computer). In terms of physical environment, the two scenarios described situations taking place at different times (morning vs. evening), different days of the
week (Thursday vs. Monday), and in different physical locations (computer lab vs. dormitory room). In terms of social/cultural environment, the Climate Change scenario asked participants to find resources to include in a PowerPoint presentation for a school assignment while the First Computer scenario asked participants to find information as part of a $20 bet with a friend.

3.2.2 Dependent Variables

Five dependent variables were used in the study. Four variables (USEFULNESS, USABILITY, AESTHETICS, and ENJOYMENT) were measured using multiple items rated on a seven-point Likert scale where 1 was “strongly disagree” and 7 was “strongly agree.” The remaining variable (GOODNESS) was measured with a single item on a seven-point semantic differential scale with “bad” and “good” as anchors. These five dependent variables will be described in this section.

3.2.2.1 USEFULNESS

The first dependent variable, and the primary variable of interest in this study, is USEFULNESS. The construct of usefulness is typically measured using a rating scale developed by Davis (1989) as part of the Technology Acceptance Model (TAM), where perceived usefulness (PU) was defined as “the degree to which a person believes that using a particular system would enhance his or her job performance” (p. 320). The original rating scale for measuring PU includes items related to effectiveness (e.g., “Using [the system] would enhance my effectiveness on the job” and “Using [the system] would improve my job performance”), and efficiency (e.g., “Using [the system] in my job would enable me to accomplish tasks more quickly” and “Using [the system] would improve my job performance”), both of which are commonly cited in the HCI literature.
as elements of usability. Due to possible confounding effects with the USABILITY variable (see Section 3.2.2.2), the TAM scale for measuring USEFULNESS was inappropriate for the current study.

Instead, the items representing USEFULNESS were developed by the researcher based on the proposed definition of usefulness (see Section 2.2) as “the extent to which a system’s functions allow users to complete a set of tasks and achieve specific goals in a particular context of use.” This definition was developed after a review of the HCI literature showed four commonly cited elements of usefulness: functions, tasks, goals, and use context. The resulting sub-scale for measuring USEFULNESS was based on this definition and consisted of the five items listed in Table 6.

<table>
<thead>
<tr>
<th>Item(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The website provides the right functions.</td>
<td>Developed by the</td>
</tr>
<tr>
<td>2. I am able to use the website to complete my task(s).</td>
<td>researcher</td>
</tr>
<tr>
<td>3. I am able to use the website to fulfill my goal(s).</td>
<td></td>
</tr>
<tr>
<td>4. The website fits my current situation.</td>
<td></td>
</tr>
<tr>
<td>5. The website is useful.</td>
<td></td>
</tr>
</tbody>
</table>

### Table 6. Items for measuring USEFULNESS.

**3.2.2.2 USABILITY**

The items representing USABILITY were originally included on the Interface Quality Scale (IQS) developed by De Angeli, Hartmann, & Sutcliffe (2009). The scale consisted of four items based on the ISO definition of usability in terms of effectiveness, efficiency, and satisfaction. The sub-scale for measuring USABILITY is listed in Table 7.
Table 7. Items for measuring USABILITY.

<table>
<thead>
<tr>
<th>Item(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The website is easy to use.</td>
<td>De Angeli, Hartmann, &amp; Sutcliffe (2009)</td>
</tr>
<tr>
<td>2. I feel in control when I am using this website.</td>
<td></td>
</tr>
<tr>
<td>3. The website requires little effort to use.</td>
<td></td>
</tr>
<tr>
<td>4. Using the website is effective.</td>
<td></td>
</tr>
</tbody>
</table>

3.2.2.3 AESTHETICS

The items representing AESTHETICS were from the Visual Aesthetics of Websites Inventory (VisAWI), an 18-item survey consisting of four main components of visual aesthetics: simplicity, diversity, colorfulness, and craftsmanship (Moshagen & Thielsch, 2010). This study employed a shortened, four-item version of the VisAWI (termed the VisAWI-4) which has been shown to have high internal consistency and high correlation with the original 18-item VisAWI (Moshagen, personal communication). A multi-item rating scale for measuring aesthetics is consistent with the finding that aesthetics is a multi-dimensional construct (Schenkman & Jönsson, 2000; Lavie & Tractinsky, 2004). The sub-scale for AESTHETICS is listed in Table 8.

Table 8. Items for measuring AESTHETICS.

<table>
<thead>
<tr>
<th>Item(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Visually, everything goes together on this website.</td>
<td>Moshagen &amp; Thielsch (2010); Moshagen (personal communication)</td>
</tr>
<tr>
<td>2. The color composition is attractive.</td>
<td></td>
</tr>
<tr>
<td>3. The layout appears professionally designed.</td>
<td></td>
</tr>
<tr>
<td>4. The layout is pleasantly varied.</td>
<td></td>
</tr>
</tbody>
</table>
3.2.2.4 ENJOYMENT

The items representing ENJOYMENT were from van Shaik and Ling (2011), who defined enjoyment as “the degree to which the use of an interactive system will be enjoyable in its own right, irrespective of any external (performance) rewards” (p. 20). The inclusion of ENJOYMENT as a dependent variable was intended to reflect the importance of considering the emotional element associated with the user experience (UX). The ENJOYMENT variable was not intended to serve as a proxy for the entire emotional aspect of systems use, which is often complex and multi-faceted (e.g., Forlizzi & Battarbee, 2004). Instead, measures of ENJOYMENT were included in recognition that system use has both pragmatic and hedonic components, and both aspects should be included when evaluating UX (Hassenzahl, 2001; Lavie & Tractinsky, 2004). The sub-scale for measuring ENJOYMENT is listed in Table 9.

Table 9. Items for measuring ENJOYMENT.

<table>
<thead>
<tr>
<th>Item(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I find using the website to be enjoyable.</td>
<td>van Shaik &amp; Ling (2011)</td>
</tr>
<tr>
<td>2. The actual process of using the website is pleasant.</td>
<td></td>
</tr>
<tr>
<td>3. I have fun using the website.</td>
<td></td>
</tr>
</tbody>
</table>

3.2.2.5 GOODNESS

The final dependent variable, GOODNESS, has been shown to influence user perceptions (Hassenzahl, 2004). It is considered representative of a user’s overall judgment or evaluation of a system, and is thus considered a higher-level construct than the other dependent variables (e.g., USEFULNESS, USABILITY). It was measured using a single item on a 7-point semantic differential scale and is presented in Table 10.
Table 10. Item for measuring GOODNESS.

<table>
<thead>
<tr>
<th>Item(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I judge the website to be: bad — good (7-point scale)</td>
<td>Hassenzahl, 2004</td>
</tr>
</tbody>
</table>

3.2.3 Summary

The variables included in this study are summarized in Table 11. They include two independent variables (CONTEXT and SYSTEM), each with three levels, and five dependent variables (USEFULNESS, USABILITY, AESTHETICS, ENJOYMENT, and GOODNESS), each measured with 7-point rating scales.

Table 11. Summary of variables used in the study.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Type</th>
<th>Levels/Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSTEM</td>
<td>Independent (3 levels)</td>
<td>3 systems: ipl2, RefSeek, Awesome Library</td>
</tr>
<tr>
<td>CONTEXT</td>
<td>Independent (3 levels)</td>
<td>3 contextual scenarios: Climate Change, First Computer, Control</td>
</tr>
<tr>
<td>USEFULNESS</td>
<td>Dependent</td>
<td>1. The website provides the right functions.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. I am able to use the website to complete my task(s).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. I am able to use the website to fulfill my goal(s).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. The website fits my current situation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. The website is useful.</td>
</tr>
<tr>
<td>USABILITY</td>
<td>Dependent</td>
<td>6. The website is easy to use.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7. I feel in control when I am using this website.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8. The website requires little effort to use.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9. Using the website is effective.</td>
</tr>
<tr>
<td>AESTHETICS</td>
<td>Dependent</td>
<td>10. Everything goes together on this website.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11. The color composition is attractive.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12. The layout appears professionally designed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13. The layout is pleasantly varied.</td>
</tr>
<tr>
<td>ENJOYMENT</td>
<td>Dependent</td>
<td>14. I find using the website to be enjoyable.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15. The actual process of using the website is pleasant.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16. I have fun using the website.</td>
</tr>
<tr>
<td>GOODNESS</td>
<td>Dependent</td>
<td>17. I judge the website to be: bad—good</td>
</tr>
</tbody>
</table>
3.3 Experimental Design

The experiment was an incomplete within-subjects experiment with two independent variables (CONTEXT and SYSTEM) and five dependent variables (USEFULNESS, USABILITY, AESTHETICS, ENJOYMENT, GOODNESS). Within-subjects designs (also called “repeated measures” designs) differ from independent groups designs (e.g., between-groups or mixed designs) in several ways. In an independent groups design, treatments are administered by randomly assigning subjects to an experimental group. Each group then receives one treatment, which serves as an experimental control because there is equal chance that subjects with “extreme” characteristics are assigned to each group. By contrast, in a within-subjects design all treatments are administered to all subjects, implying that each subject serves as their own experimental control. There are two types of within-subjects designs: complete within-subjects designs, where all subjects are exposed to all treatment conditions, and incomplete within-subjects designs (like the current study), where all subjects are exposed to each treatment condition exactly once (Shaughnessy, Zechmeister, & Zechmeister, 2012).

There are three advantages to using a within-subjects design. First, fewer subjects are necessary in within-subjects designs because researchers can collect multiple measurements from each subject. This allows for faster and more efficient subject recruitment, particularly when the population of potential subjects is limited. Second, within-subjects experiments are more sensitive than independent group designs because there is less variation in the way subjects respond to each treatment condition; that is, the same subject is likely to respond similarly to two different treatments whereas two
different subjects may respond differently to different treatments. This leads to less error variance and a greater chance of detecting the effects of an independent variable. Third, because all subjects are exposed to all treatment conditions, the effects of independent variables cannot be confounded with individual differences; that is, one treatment condition cannot end up with happier, smarter, or healthier subjects than the other treatment conditions. This means researchers can be confident that any effects are due to changes in the independent variables rather than to differences in the subject populations (Ryan, 2007; Shaughnessy, et al., 2012).

The disadvantage of within-subjects designs is that because multiple measurements are taken on the same subject, there is a possibility that subjects get better (with increased practice) or worse (due to boredom, fatigue, or lack of motivation) at completing the experimental tasks. These effects are called practice effects (also called learning effects or carryover effects) and their impact can be minimized through the process of counterbalancing, which is different for complete and incomplete designs. In complete designs, treatments are counterbalanced within subjects by administering every treatment to each subject multiple times. In incomplete designs, treatments are counterbalanced across subjects by administering each treatment only once to each subject and ensuring that each treatment occurs in each ordinal position an equal number of times across all subjects (Shaughnessy, et al., 2012).

The current study is an incomplete within-subjects design with two independent variables, each with 3 levels (SYSTEM and CONTEXT). Both independent variables are completely within-subjects variables, meaning that each subject was asked to interact with all three levels of the SYSTEM variable and all three levels of the CONTEXT
variable. Studies have shown that repeated use of a system or task improves user performance (e.g., Grossman, et al., 2009). To minimize these effects, each subject was asked to interact with each SYSTEM exactly once and each CONTEXT exactly once, thus making the experiment an incomplete within-subjects design. This was achieved by creating six sets of three treatments, where each treatment consisted of a SYSTEM and CONTEXT pair (see Figure 7).

The sequence of treatments was counterbalanced within each set so that every treatment appeared in each ordinal position an equal number of times. The three possible treatments created six ($3! = 3 \times 2 \times 1 = 6$) possible sequences for each set. Therefore, a completely counterbalanced incomplete within-subjects design was created from 36 different sequences of treatments (6 sets x 6 sequences per set).

3.4 Experimental Procedure

The experiment was conducted in a usability lab at the iSchool at Drexel University. Participants met individually with the researcher to complete the experiment.
At the beginning of the experiment, participants were randomly assigned to one of the six sets and presented with three treatments, one at a time. Each treatment combination consisted of a contextual scenario (CONTEXT) and a website (SYSTEM). For each treatment, participants were given five minutes to think aloud while attempting to complete the scenario on the given system. At the end of five minutes, they were asked to think aloud while completing a post-scenario survey consisting of the 17 items measuring the dependent variables of USEFULNESS, USABILITY, AESTHETICS, ENJOYMENT, and GOODNESS. After all three treatments were administered, participants were asked to complete a post-study questionnaire. The entire study took between 30 and 35 minutes to complete.

This study utilized a traditional think aloud protocol in which participants verbalized their thought process while interacting with each website, with no interventions or interruptions from the researcher. Although users do not typically think aloud when interacting with websites in the real world, it is a widely used method in usability research (McDonald, Edwards, & Zhao, 2012). While there is the possibility that using think aloud influences participant behavior, Ericsson and Simon (1993) argued that talking aloud does not interfere with task performance. This assertion was confirmed by a recent study of think aloud protocols which showed that there were no significant differences in task performance between the think aloud and non-think aloud conditions (Olmsted-Hawala, et al., 2010). Therefore, while verbalizing their thoughts may have made participants more self-aware or self-conscious, it is unlikely that the think aloud process had a significant influence on the results of the experiment. And, although it is beyond the scope of the current dissertation, a secondary analysis of the think aloud data
may reveal further insight into how participants experience, perceive and evaluate websites.

3.5 Manipulation Check

A manipulation check assesses whether the experimental manipulation (in this case, the level of CONTEXT) actually caused a discernible difference in participants' behavior. An informal review of the think aloud recordings showed that all 36 participants explicitly discussed the specific contextual scenario (and did not mention any other scenarios) when filling out the post-scenario survey. Additionally, steps were taken to ensure that participants’ responses on the post-scenario survey were directly related to the contextual scenario they experienced. First, participants were not permitted to interact with the website until the researcher was certain they had a firm understanding of what they were supposed to do. To accomplish this, participants were asked to read the contextual scenario aloud and briefly summarize it. In the few occasions where it seemed unclear whether participants understood the details of the scenario, the researcher asked a series of follow-up questions until they exhibited a firm understanding. Second, under the control (no scenario) condition, participants were instructed to explore the website in any way they saw fit. But, if any participants expressed interest in using search tasks from the other two scenarios, the researcher directed them to think of alternative tasks so that the three contextual scenarios remained separate and distinct interactive experiences. Consequently, there is strong evidence that the manipulation was successful.

3.6 Data Collection Instruments

There were two data collection instruments utilized in this study: a post-scenario survey and a post-study questionnaire. Data from both instruments were collected using
SurveyGizmo\textsuperscript{8}, an online survey creation and management tool. The items from the post-scenario survey were presented in random order for each participant to reduce order effects, with the exception of the item representing GOODNESS which was always presented last (since it is assumed to represent a user’s overall evaluation).

\subsection*{3.6.1 Post-Scenario Survey}

The post-scenario survey consisted of the 17 items measuring the five dependent variables: 16 items measured USEFULNESS (5 items), USABILITY (4 items), AESTHETICS (4 items), and ENJOYMENT (3 items) with a seven-point Likert scale where 1 was “strongly disagree” and 7 was “strongly agree”, and one item measured GOODNESS using a seven-point semantic differential scale with “bad” and “good” as anchors. A seven-point scale was chosen because five-point scales have been shown to lack sensitivity in many cases (Finstad, 2010a; Finstad, 2010b). In addition, Cox (1980) determined that seven was the optimal number of alternatives for Likert scales and other studies have shown that seven-point scales outperform five-point scales in terms of reliability, accuracy and ease of use (Diefenbach, Weinstein, & O'Reilly, 1993). The items from the post-scenario survey are described in Section 3.2.2. The post-scenario survey is presented in Appendix A. The first step in data analysis was to perform a confirmatory factor analysis (CFA), which is a test of how well a set of rating sub-scales actually measure the constructs they were intended to measure (Hair, et al., 2010). A CFA is used to a) determine the goodness of fit of a hypothesized model, b) confirm the construct validity of a measurement instrument, and c) calculate factor scores for each construct. These three aspects of CFA will be explained in this section.

\footnote{\texttt{http://www.surveygizmo.com/}}
3.6.1.1 Goodness of Fit

The first goal of CFA is to determine how well a hypothesized model fits a set of data. According to Hair et al. (2010), it is good practice in CFA to include at least three items to measure each latent construct. Therefore, the CFA was used to determine whether four dependent variables (USEFULNESS, USABILITY, AESTHETICS, and ENJOYMENT) were actually measured by the four sub-scales listed in Section 3.2.2. The dependent variable GOODNESS was not included in the CFA because it only contained a single item on its sub-scale. The CFA was conducted in SPSS AMOS, a structural equation modeling (SEM) tool that allows researchers to evaluate how well a given data set fits a hypothesized model.

To evaluate a given model, SPSS AMOS produces several “goodness of fit” indices and a model is considered adequate if multiple indices fall within an accepted range. The most commonly used goodness of fit indices were identified by Sun (2005), Schreiber et al. (2006), Hooper, Coughlan, and Mullen (2008), and Mueller and Hancock (2008). They are listed in Table 12 with their definitions and acceptance criteria.
### Table 12. Goodness of fit indices for confirmatory factor analysis.

<table>
<thead>
<tr>
<th>Index</th>
<th>Definition</th>
<th>Acceptance Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>Model Chi-square</td>
<td>$p &gt; 0.05^9$</td>
</tr>
<tr>
<td>SRMR</td>
<td>Standardized Root Mean Square Residual</td>
<td>$\leq 0.08$</td>
</tr>
<tr>
<td>RMSEA</td>
<td>Root Mean Square Error of Approximation</td>
<td>$\leq 0.06$ to $0.08$ (with confidence interval)</td>
</tr>
<tr>
<td>CFI</td>
<td>Comparative Fit Index</td>
<td>$\geq 0.95$</td>
</tr>
<tr>
<td>TLI</td>
<td>Tucker-Lewis Index</td>
<td>$\geq 0.95$</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
<td>Used for comparing models</td>
</tr>
</tbody>
</table>

All sub-scales used on the post-scenario survey, with the exception of the USEFULNESS sub-scale, were adapted from previous studies and known to be valid and reliable indicators of the underlying latent variables. However, none of the three validated sub-scales had been used together in previous studies, and one sub-scale (USEFULNESS) was developed specifically for this study. The hypothesized full model is shown graphically in Figure 8.

---

$^9$ The acceptance criterion for the $\chi^2$ test is a non-significant result ($p > 0.05$) because the null hypothesis is that the model is a perfect fit for the data. Thus, a significant result indicates that the model is a poor fit for the data while a non-significant result indicates that the model is a good fit (Sun, 2005).
In the figure above, the four dependent variables (USEFULNESS, USABILITY, AESTHETICS, and ENJOYMENT) are represented by the ovals on the right; the arrows between the four dependent variables indicate that the four variables are likely correlated with each other. The measurement items for each sub-scale are represented in boxes, and each box is connected by an arrow to the construct it was intended to measure. To the left of each measurement item is a circle representing the error term for each item.

Running the CFA in SPSS AMOS for the full model produced the following goodness of fit indices: $\chi^2 (98) = 266.60$, $p < 0.001$; RMSEA = 0.13 with 90% C.I. of
0.11 to 0.15; SRMR = 0.07; CFI = 0.90 and TLI = 0.88. In this case, none of the
goodness of fit indices met the acceptance criteria listed in Table 12, leading to a
rejection of the model. This result suggests that the 16 items from the post-scenario
survey do not adequately measure the four constructs of USEFULNESS, USABILITY,
AESTHETICS, and ENJOYMENT and it is necessary to re-specify the model.

A common method of re-specification is examining the modification indices. A
modification index is calculated for every relationship not estimated in the model (i.e.,
the relationships between the latent constructs and the error terms for each measurement
item) and represents the reduction in the $\chi^2$ statistic if the given relationship was
estimated in the model. A modification index over 4.0 suggests that the model could be
improved if the relationship were included in the model. However, relationships should
only be added if they make sense from a theoretical perspective. For instance, the
relationship between two error terms from different constructs should not be added to the
model because it could indicate a problem with discriminant validity. In these cases, the
researcher may decide to remove the offending items or reduce the number of constructs
included in the model (Hair, et al., 2010). The goal of the CFA is to re-specify the model
until either the model is acceptable according to the goodness of fit indices or no further
theoretical modifications can be made to the model.

According to the output produced by SPSS AMOS, 30 unobserved relationships
had a modification index over 4.0, indicating poor model fit (Hair, et al., 2010). A closer
examination of these modification indices showed that three measurement items appeared
to be strongly correlated with multiple constructs. The problematic items were $e_{15}$ (“The
actual process of using the website is pleasant”), $e_{9}$ (“Using the website is effective”)
and e1 ("The website provides the right functions"). The modification indices relevant to these three items are presented in Table 13.

<table>
<thead>
<tr>
<th>First Item</th>
<th>Second Item</th>
<th>Modification Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>e14</td>
<td>e16</td>
<td>38.797</td>
</tr>
<tr>
<td>e2</td>
<td>e3</td>
<td>35.495</td>
</tr>
<tr>
<td>e9</td>
<td>USEFULNESS</td>
<td>23.455</td>
</tr>
<tr>
<td>e3</td>
<td>e1</td>
<td>13.344</td>
</tr>
<tr>
<td>e15</td>
<td>USABILITY</td>
<td>11.086</td>
</tr>
<tr>
<td>e6</td>
<td>e1</td>
<td>6.275</td>
</tr>
<tr>
<td>e15</td>
<td>e6</td>
<td>5.430</td>
</tr>
<tr>
<td>e1</td>
<td>USABILITY</td>
<td>5.348</td>
</tr>
<tr>
<td>e12</td>
<td>e1</td>
<td>4.966</td>
</tr>
<tr>
<td>e15</td>
<td>e7</td>
<td>4.834</td>
</tr>
<tr>
<td>e10</td>
<td>e1</td>
<td>4.529</td>
</tr>
</tbody>
</table>

From Table 13, e9 has a strong correlation with USEFULNESS, e15 has a strong correlation with USABILITY, and e1 has a strong correlation with e6, USEABILITY, e12 and e10, suggesting that the item “The website provides the right functions” is related to USEFULNESS, USABILITY, and AESTHETICS. Based on these results, it is decided that three items should be dropped from the model: “Using the website is effective” (e9), “The actual process of using the website is pleasant” (e15), and “The website provides the right functions” (e1). Some items from the same subscales are also highly correlated. Based on these findings, an estimate of the relationship between e2 ("I am able to use the website to complete my tasks") and e3 ("I am able to use the website to complete my goals") should be added to the model. An estimate of the
relationship between $e_{14}$ ("I find using the website to be enjoyable") and $e_{16}$ ("I have fun using the website") does not need to be added since their relationship will already be included due to the elimination of $e_{15}$ ("The actual process of using the website is pleasant"). The revised model is presented graphically in Figure 9.
Running the CFA in SPSS AMOS for the revised model produced the following goodness of fit indices: $\chi^2 (58) = 75.71$, $p = 0.059$; RMSEA = 0.05 with 90% C.I. of 0.00 to 0.08; SRMR = 0.04; CFI = 0.99 and TLI = 0.98. For the revised model, all of the goodness of fit indices met the acceptance criteria listed in Table 12, leading to acceptance of the revised model. This result suggests that 13 items from the post-scenario survey are adequate measures of the four constructs. Now that the model goodness of fit has been established, it is necessary to examine the construct validity of the post-scenario survey.
3.6.1.2 Construct Validity of the Post-Scenario Survey

Construct validity refers to the extent to which a set of items actually reflect the latent constructs they were intended to measure. It consists of four types of validity: convergent validity, discriminant validity, nomological validity, and face validity (Hair, et al., 2010). Face and nomological validity are typically assessed qualitatively while convergent and discriminant validity are assessed quantitatively. Violations to construct validity suggest that an instrument does not adequately measure the constructs in question because either the items are poor measures of the constructs (convergent validity), the items cannot distinguish between different constructs (discriminant validity), the items do not make sense as indicators of a construct (face validity) or the constructs have unreasonable or unexpected correlations with each other (nomological validity). Thus, an acceptable goodness of fit is not sufficient for determining an adequate model; it must also exhibit construct validity. This section will present and examine the four types of construct validity for the post-scenario survey.

Convergent Validity. Convergent validity, which refers to whether a set of variables measure the same construct, is achieved when all standardized factor loadings of each item on their associated construct exceed 0.70, the average variance extracted (AVE) for each construct exceeds 0.50, and the Cronbach’s alpha for each construct exceeds 0.70 (Hair, et al., 2010). Measures of convergent validity and their acceptance criteria are summarized in Table 14.
Table 14. Measures and acceptance criteria for convergent validity.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Acceptance Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual item reliability</td>
<td>Standardized factor loadings for each item in a construct exceed 0.70</td>
</tr>
<tr>
<td>Average Variance Extracted (AVE)</td>
<td>AVE for each construct exceeds 0.50</td>
</tr>
<tr>
<td>Construct reliability (CR)</td>
<td>Cronbach’s alpha for each factor exceeds 0.70</td>
</tr>
</tbody>
</table>

The convergent validity of the revised model (Model 2) is confirmed because all factor loadings exceed 0.70, the AVE for each construct exceeds 0.50, and Cronbach’s alpha for each construct exceeds 0.70. These results are summarized in Table 15 below.
Table 15. Convergent validity of the revised model (Model 2)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Measurement Item</th>
<th>Std. Factor Loading</th>
<th>AVE</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>USEFULNESS</td>
<td>1. I am able to use the website to complete my task(s).</td>
<td>0.866</td>
<td>0.926</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>2. I am able to use the website to fulfill my goal(s).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. The website fits my current situation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. The website is useful.</td>
<td>0.851</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USABILITY</td>
<td>5. The website is easy to use.</td>
<td>0.865</td>
<td>0.710</td>
<td>0.879</td>
</tr>
<tr>
<td></td>
<td>6. I feel in control when I am using this website.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. The website requires little effort to use.</td>
<td>0.836</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AESTHETICS</td>
<td>8. Visually, everything goes together on this website.</td>
<td>0.826</td>
<td>0.672</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>9. The color composition is attractive.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10. The layout appears professionally designed.</td>
<td>0.816</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11. The layout is pleasantly varied.</td>
<td>0.809</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENJOYMENT</td>
<td>12. I find using the website to be enjoyable.</td>
<td>0.995</td>
<td>0.855</td>
<td>0.915</td>
</tr>
<tr>
<td></td>
<td>13. I have fun using the website.</td>
<td>0.849</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Discriminant Validity.** Discriminant validity, which refers to the extent to which the constructs are truly distinct from each other, is achieved when the factor loadings of items on their associated construct are higher than the cross-loadings of items from other constructs and when the AVE of each construct is greater than the squared correlations between the constructs (Hair, et al., 2010). Measures of discriminant validity and their acceptance criteria are listed in Table 16.
Table 16. Measures and acceptance criteria for discriminant validity.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Acceptance Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor loadings and cross-loadings</td>
<td>Factor loadings for each item in a construct are higher than cross-loading of items from other constructs</td>
</tr>
<tr>
<td>AVE and squared correlations</td>
<td>AVE for each construct is higher than the squared correlations between constructs</td>
</tr>
</tbody>
</table>

The factor loadings of each item on each of the four constructs are presented in Table 17. The highest factor loading of each item is on its associated factor (rows) and the items with the highest factor loadings for each construct are the intended items (columns).
Table 17. Factor loadings of measurement items on the four constructs.

<table>
<thead>
<tr>
<th>Measurement Item</th>
<th>USEFUL</th>
<th>USABILITY</th>
<th>AESTH</th>
<th>ENJOY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I am able to use the website to complete my task(s).</td>
<td>0.131</td>
<td>0.022</td>
<td>-0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>2. I am able to use the website to fulfill my goal(s).</td>
<td>0.091</td>
<td>0.016</td>
<td>-0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>3. The website fits my current situation.</td>
<td>0.163</td>
<td>0.028</td>
<td>-0.004</td>
<td>0.001</td>
</tr>
<tr>
<td>4. The website is useful.</td>
<td>0.423</td>
<td>0.072</td>
<td>-0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>5. The website is easy to use.</td>
<td>0.044</td>
<td>0.235</td>
<td>0.040</td>
<td>0.004</td>
</tr>
<tr>
<td>6. I feel in control when I am using this website.</td>
<td>0.038</td>
<td>0.202</td>
<td>0.035</td>
<td>0.003</td>
</tr>
<tr>
<td>7. The website requires little effort to use.</td>
<td>0.032</td>
<td>0.169</td>
<td>0.029</td>
<td>0.003</td>
</tr>
<tr>
<td>8. Visually, everything goes together on this website.</td>
<td>-0.005</td>
<td>0.031</td>
<td>0.198</td>
<td>0.001</td>
</tr>
<tr>
<td>9. The color composition is attractive.</td>
<td>-0.005</td>
<td>0.031</td>
<td>0.198</td>
<td>0.001</td>
</tr>
<tr>
<td>10. The layout appears professionally designed.</td>
<td>-0.004</td>
<td>0.026</td>
<td>0.169</td>
<td>0.001</td>
</tr>
<tr>
<td>11. The layout is pleasantly varied.</td>
<td>-0.004</td>
<td>0.029</td>
<td>0.185</td>
<td>0.001</td>
</tr>
<tr>
<td>12. I find using the website to be enjoyable.</td>
<td>0.071</td>
<td>0.109</td>
<td>0.044</td>
<td>0.760</td>
</tr>
<tr>
<td>13. I have fun using the website.</td>
<td>0.002</td>
<td>0.004</td>
<td>0.001</td>
<td>0.025</td>
</tr>
</tbody>
</table>

The AVE (average variance extracted) and squared correlations for each construct are presented in Table 18. For each construct, the AVE is higher than the squared correlations between itself and any other construct, confirming the discriminant validity.
Table 18. Squared multiple correlations for each construct.

<table>
<thead>
<tr>
<th></th>
<th>AVE</th>
<th>Usefulness</th>
<th>Usability</th>
<th>Aesthetics</th>
<th>Enjoyment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>0.926</td>
<td>0.671</td>
<td>0.272</td>
<td>0.549</td>
<td>0.549</td>
</tr>
<tr>
<td>Usability</td>
<td>0.710</td>
<td>0.671</td>
<td>-</td>
<td>0.546</td>
<td>0.645</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>0.672</td>
<td>0.272</td>
<td>0.546</td>
<td>-</td>
<td>0.392</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>0.855</td>
<td>0.549</td>
<td>0.645</td>
<td>0.392</td>
<td>-</td>
</tr>
</tbody>
</table>

Nomological Validity. Nomological validity, which is a post-hoc assessment of whether the correlations between the constructs make sense from a theoretical standpoint, is assessed by examining the correlations between the constructs and making a subjective determination about whether these correlations make sense. For the revised model, all four constructs had a statistically significant correlation ($p < 0.001$) with each other (see Table 19). Since high values for each construct are associated with positive system evaluations, the statistically significant correlations between the constructs make sense from a theoretical perspective. For instance, a number of researchers have already identified a correlation between aesthetics and usability (e.g., Tractinsky, 1997). Thus, nomological validity of the measurement instrument is confirmed.
Table 19. Correlations between the four dependent variables.

<table>
<thead>
<tr>
<th></th>
<th>Usefulness</th>
<th>Usability</th>
<th>Aesthetics</th>
<th>Enjoyment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>-</td>
<td>0.819</td>
<td>0.522</td>
<td>0.741</td>
</tr>
<tr>
<td>Usability</td>
<td>0.819</td>
<td>-</td>
<td>0.739</td>
<td>0.803</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>0.522</td>
<td>0.739</td>
<td>-</td>
<td>0.626</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>0.741</td>
<td>0.803</td>
<td>0.626</td>
<td>-</td>
</tr>
</tbody>
</table>

**Face Validity.** Face validity, which is established *a priori* and refers to whether the item measurements make sense as indicators of a latent variable, is assessed by examining the content of the measurement items for each sub-scale. As mentioned previously, three of the sub-scales used in this study were adapted from prior research. The AESTHETICS sub-scale was from the Visual Aesthetics of Websites Inventory (VisAWI), an 18-item survey consisting of four main components of visual aesthetics: simplicity, diversity, colorfulness, and craftsmanship (Moshagen & Thielsch, 2010). This study employed a four-item version of the VisAWI (the VisAWI-4) which was shown to have high internal consistency and high correlation with the 18-item VisAWI (Moshagen, personal communication). The ENJOYMENT sub-scale was from van Shaik and Ling (2011), who define it as “the degree to which the use of an interactive system will be enjoyable in its own right, irrespective of any external (performance) rewards” (p. 20). The USABILITY sub-scale was from the Interface Quality Scale (IQS) developed by De Angeli, Hartmann, & Sutcliffe (2009) and was based on the ISO definition of usability. The USEFULNESS sub-scale was developed by the author based on a definition of
usefulness as “the extent to which a system’s functions allow users to complete a set of
tasks and achieve specific goals in a particular context of use.” This definition was
proposed after an extensive review of the HCI research literature indicated that
researchers employ the term usefulness (or utility) in terms of system functions, users’
tasks, users’ goals, or the users’ context or situation. Even though the revised model does
not include the item regarding the system functions, the sub-scale remains a reasonable
measurement of the usefulness construct. Since there is sound theoretical justification for
using all four sub-scales, the face validity of the measurement instrument is confirmed.

3.6.1.3 Factor Scores for the Dependent Variables

Once a model demonstrates both good fit and construct validity, factor scores for
the unobserved constructs can be calculated. Unlike composite scores which are typically
calculated by taking the sum or mean of the items of each sub-scale, factor scores are
calculated using the matrix of factor loadings. Thus, factor scores are more informative
than composite scores because they consider the entire factor structure of the
measurement instrument rather than just the item scores of each individual sub-scale.

Factor scores for the dependent variables in this study were automatically calculated in
SPSS AMOS. Since the revised model was acceptable based on the goodness of fit
indices and the construct validity of the measurement instrument was confirmed, factor
scores for all four constructs were calculated for each of the 108 observations using the
matrix of factor loadings (Table 17). From this point forward, references to
USEFULNESS, USABILITY, AESTHETICS, or ENJOYMENT will refer to these factor
scores.
3.6.1.4 Summary

A confirmatory factor analysis of the full model including 16 items from the post-scenario survey measuring USEFULNESS, USABILITY, AESTHETICS, and ENJOYMENT did not achieve adequate goodness of fit. Further analysis showed that three items were highly correlated with other constructs, leading to their removal from the model. The revised model successfully achieved goodness of fit and was also shown to have construct validity in terms of convergent, discriminant, nomological, and face validity, suggesting that the revised model provides valid measures of the four constructs. Therefore, factor scores were calculated for each of construct using participants’ ratings on the remaining 13 items from the post-scenario survey.

3.6.2 Post-Study Questionnaire

At the end of the experiment, participants completed a post-study questionnaire. First, participants were asked respond verbally to the following question regarding their definition of the term usefulness:

*What makes a website useful to you? What criteria do you look for?*

Next, participants were asked whether they have experience using the systems utilized in this study. Next, participants were asked to provide their knowledge of the Internet, which was assessed using the Internet Knowledge Measure (iKnow). The iKnow is composed of 14 items, measured on a six-point Likert-type scale, regarding common terms, knowledge and behaviors associated with the Internet. Participants’ iKnow scores are calculated by taking the sum of their answers on all 14 items of the iKnow questionnaire (Potosky, 2007). Finally, participants were asked to provide their basic demographic information (age, gender) and background (major, class standing, number of HCI courses taken). The post-study questionnaire is presented in Appendix B.
3.7 Data Analysis Methods

This section will describe the data analysis methods used in this study and described in Chapter 4. First, the overall approach to data analysis will be described. Next, the primary quantitative data analysis method (marginal models) will be explained. Finally, the constant comparative method of qualitative analysis will be described.

3.7.1 Approach to Data Analysis

Data collected from the post-scenario survey on the five dependent variables (USEFULNESS, USABILITY, AESTHETICS, ENJOYMENT, and GOODNESS) were in the form of ratings on 7-point Likert scales. Some researchers believe that data collected from Likert scales violate basic assumptions required for parametric data analysis, resulting in inflated Type I error rates (Kaptein, et al., 2010; Wobbrock, et al., 2011). However, others have argued that these criticisms are unfounded because parametric methods are actually robust to violations of these assumptions (Norman, 2010). Since there seems to be no consensus, parametric analysis methods were selected for this study.

3.7.2 Marginal Models

This section will describe marginal models. First, marginal models will be described in terms of their advantages and appropriateness for the current study. Next, the model building approach for marginal models will be explained.

3.7.2.1 Overview of Marginal Models

The current study adopted an incomplete within-subjects experimental design, meaning that repeated measures were taken on each participant. In this case, participants were asked to complete the post-scenario survey three times, once for each treatment
(SYSTEM x CONTEXT pair). Because of the within-subjects design, it is not reasonable to assume that the observations are independent from each other; that is, participants’ ratings of the three systems are likely to be related. In such cases, Hassenzahl and Monk (2010) recommended creating two separate datasets by aggregating ratings across systems and across participants. Using this approach, both datasets are analyzed separately and results are considered valid if they are similar for each dataset. However, in the current study it was expected that participants would provide different ratings depending on the SYSTEM and CONTEXT. If data were aggregated by participant, critical data would be lost about the effects of these independent variables. Additionally, aggregating by system would yield insufficient data because only three systems were used in this study. For these reasons, using either data aggregation strategy was not feasible. Instead, the data were analyzed using a marginal model.

Marginal models can be considered a subset of Linear Mixed Models (LMMs), which are statistical models in which the residuals are not assumed to be independent or to have constant variance, making them popular in repeated-measures or longitudinal studies in which subjects are measured multiple times under different conditions (West, Welch, & Gatecki, 2007). Because measurements taken on the same variable for the same subject are likely to be correlated, models fitted to repeated-measures data require estimation of the covariance parameters that explain this correlation. This estimation is typically performed with repeated-measures ANOVA models, but LMMs offer two important advantages. First, a traditional repeated-measures ANOVA model requires estimation of all covariance parameters or an assumption of “sphericity” of the covariance matrix. LMMs allow for many different covariance structures, giving
researchers more flexibility and efficiency when estimating the relationship between repeated measurements. Second, traditional repeated-measures ANOVA models perform a “complete-case analysis” in which subjects are dropped from the model entirely if any measurements are missing (West, et al., 2007). LMMs do not have this restriction, making them appropriate when all subjects are not measured under all possible conditions (like the current study). In the current study, models were created using the Mixed Models (MIXED) procedure in SPSS with REML\textsuperscript{10} estimation.

Marginal models are sometimes called “population averaged” models because they differ from LMMs in one critical way: they contain no random effects. Random effects are values assumed to be randomly sampled from a population and can take the form of random intercepts, which represent deviations from a fixed intercept for a specific subject, or random coefficients, which represent deviations for a given subject from the overall fixed effects (West, et al., 2007). Random effects are not of interest in the current study for two reasons. First, random effects models (sometimes called “subject-specific” models) aim to model the effects of individual participants on the outcome variable. By contrast, the interest of the current study was to model the effects of the fixed effects (CONTEXT and SYSTEM) on the population as a whole. Second, the study design indicated that participants were only exposed to each level of CONTEXT and SYSTEM once, meaning no repeated measurements were taken at any level of either factor. Since participants were exposed to a different treatment for each repeated measure, the random effects of the treatments are confounded with the repeated

\textsuperscript{10} REML refers to the method used to obtain the unknown parameters in LMMs. Often called “restricted maximum likelihood estimation,” REML is an alternative to Maximum Likelihood (ML) estimation that produces unbiased estimates of the covariance parameters by taking the degrees of freedom into account (West, et al., 2007).
measurements themselves, meaning that calculating random effects and repeated effects would be redundant. Therefore, it was decided to exclude random effects and only include residual variation resulting from the repeated measurements.

### 3.7.2.2 Model Building Approach

The model building approach used in the study is a top-down approach in which all possible fixed effects are added to the model and non-significant effects are removed, one at a time, until only significant effects remain (West, et al., 2007). For building LMMs, West et al. outline a four-step strategy for top-down model building:

**Step 1: Fit a model with a “loaded” mean structure.** This step requires specifying a model with all possible fixed effects and covariates and a hypothesized covariance structure. For the current study, all possible fixed effects and covariates are included in the initial model along with a repeated measurement effect with an unstructured covariance. An unstructured covariance for the repeated measurements implies that each measurement has a heterogeneous variance and heterogeneous correlations with the other measurements.

**Step 2: Select a structure for the random effects.** This step was omitted in the current study because random effects were excluded from all models.

**Step 3: Select a structure for the residuals (repeated effects).** This step involves selecting an appropriate covariance structure for the repeated measurements. In addition to the unstructured covariance structure, six common covariance structures were considered for each model: compound symmetry, compound symmetry with heterogeneous variances, first order autoregressive, first order autoregressive with heterogeneous variances, Toeplitz, and Toeplitz with heterogeneous variances (Littell,
Pendergast, & Natarajan, 2000). Each covariance structure is associated with a specific number of estimated covariance parameters. An unstructured covariance structure assumes heterogeneous variances and heterogeneous covariance, making it the least restrictive covariance structure. For example, an unstructured covariance structure for three repeated measurements requires estimation of six parameters: the variances for all three measurements and the covariance between the first and second, first and third, and second and third measurements. The seven possible covariance structures, their associated parameters, and a brief description are presented in Table 20.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Abbrev.</th>
<th>Par.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured</td>
<td>UN</td>
<td>6</td>
<td>Heterogeneous variance and heterogeneous covariance</td>
</tr>
<tr>
<td>First order autoregressive</td>
<td>AR(1)</td>
<td>2</td>
<td>Homogeneous variance; higher covariance between adjacent measurements</td>
</tr>
<tr>
<td>First order autoregressive (Heterogeneous)</td>
<td>AR(1)-H</td>
<td>4</td>
<td>Same as AR(1) with heterogeneous variances</td>
</tr>
<tr>
<td>Compound Symmetry</td>
<td>CS</td>
<td>2</td>
<td>Constant variance and constant covariance</td>
</tr>
<tr>
<td>Compound Symmetry (Heterogeneous)</td>
<td>CS-H</td>
<td>4</td>
<td>Same as CS but with heterogeneous variance</td>
</tr>
<tr>
<td>Toeplitz</td>
<td>T</td>
<td>3</td>
<td>Homogeneous variance but heterogeneous covariance</td>
</tr>
<tr>
<td>Toeplitz (Heterogenous)</td>
<td>T-H</td>
<td>5</td>
<td>Same as T but with heterogeneous variance</td>
</tr>
</tbody>
</table>

To compare the different covariance structures, a likelihood ratio test (LRT) was used. In an LRT, the -2 Log Likelihood is calculated for models with each covariance structure and compared to the -2 Log Likelihood for a model with unstructured covariance (the initial model). The difference between the two values follows a $\chi^2$ distribution with the degrees of freedom calculated as the difference between the number
of covariance parameters. A non-significant p-value indicates that the two models have approximately equal explanatory power; in these cases, it is suggested that the more parsimonious (i.e., the model with fewer parameters) is preferred (West, et al., 2007). The models can also be assessed using the Bayesian Information Criterion (BIC), where the smaller BIC is associated with a better model (Seltman, 2011).

**Step 4: Reduce the model.** This step involves eliminating non-significant fixed effects. To do so, the fixed effect with the highest p-value is identified and a new model is created with that fixed effect excluded. This process is repeated until all non-significant (p > 0.05) fixed effects are gone from the model.

### 3.7.3 Constant Comparative Method

The constant comparative method was used to analyze the qualitative data collected in this study. First, audio transcriptions were created containing answers from all 36 participants to the open-ended question regarding participants’ definitions of usefulness. Next, the transcripts were split into data points by identifying and separating each coherent thought and concept embedded within each participant’s answer. Finally, the constant comparative method (Glaser & Strauss, 1967; Lincoln & Guba, 1985) was used to analyze the data. The most common method for analyzing qualitative data (Agosto & Hughes-Hassell, 2005), the constant comparative involves iteratively reading through a transcript of verbal data and grouping and re-grouping data into categories until a coherent coding scheme is created that addresses the research questions. An analysis of the resulting codes typically leads to the creation of a typology, model, analogy or some other way of presenting the results (Agosto, et al., 2010).
3.8 Summary

This section described the methods used in this study. First, two independent variables were identified: SYSTEM, which consisted of three information portal websites, and CONTEXT, which consisted of three contextual scenarios. Five dependent variables were also identified: USEFULNESS, USABILITY, AESTHETICS, and ENJOYMENT were measured using multi-item rating scales and GOODNESS was measured with a single item. Second, an incomplete repeated measures experiment was described in which 36 participants were given three treatments – consisting of one SYSTEM and one CONTEXT – and then asked to complete a post-scenario survey in which they rated each SYSTEM on all five dependent variables. A confirmatory factor analysis (CFA) showed that the post-scenario survey provided a valid measurement of the dependent variables. At the end of the experiment, participants completed a post-study questionnaire regarding their experience with the three systems used in the study, their Internet Knowledge (iKNOW) score, which measures an individuals’ self-reported knowledge and competency with Internet technologies (Potosky, 2007), and their background and demographic information. The post-study questionnaire also included one open-ended question in which participants were asked to (verbally) define a useful website or system. Finally, the methods were explained for quantitative (marginal models) and qualitative (constant comparative method) data analysis.
4. RESULTS

This chapter reports the results of the quantitative and qualitative analysis methods used in this study. First, the study’s motivation and methods are summarized and the research questions are stated. Second, an overview of the 36 participants recruited for the study is presented. Next, results of the quantitative analysis is presented by describing the creation of two marginal models – a subset of linear mixed models that adjust for repeated measurements but contain no random effects – to determine (1) how usefulness is related to context, usability, aesthetics, and enjoyment and (2) how those constructs influence users’ overall evaluation of a website (e.g., goodness). Finally, the results of the qualitative analysis are presented by describing the themes that emerged from examining participants’ descriptions of how they would define or describe a useful website.

4.1 Introduction

Although it has implicitly been an integral part of system evaluation research, the study of usefulness has been largely ignored in HCI. This dissertation presents the results of a repeated measures experiment designed to fill this gap by investigating how usefulness is influenced by changes to the context in which a system is used and how usefulness is related to other attributes of UX, such as usability, aesthetics, and enjoyment. Two data collection instruments were used in this experiment: a post-scenario survey and a post-study questionnaire. Data from both instruments were collected using SurveyGizmo, an online survey creation and management tool. Data from the post-scenario survey consisted of 108 total observations from 36 participants (3 observations per participant). Each observation consisted of the following: SYSTEM (categorical
variable with 3 categories), CONTEXT (categorical variable with 3 categories), and participants’ ratings of USEFULNESS (4 items), USABILITY (4 items), AESTHETICS (4 items), ENJOYMENT (3 items), and GOODNESS (1 item), all measured on 7-point scales. Data from the post-study questionnaire consisted of 36 observations (1 observation per participant) and included participants’ experience with the three systems used in the study, their Internet Knowledge (iKNOW) score, which measures an individuals’ self-reported knowledge and competency with Internet technologies (Potosky, 2007), and their background and demographic information. The post-study questionnaire also included one open-ended question in which participants were asked to (verbally) define a useful website or system.

The data above were collected to answer the following four research questions:

**RQ1:** How and to what extent are perceptions of usefulness shaped by the context –the user, task, and environment– in which the system is used?

**RQ2:** How and to what extent do perceptions of usability, aesthetics, and enjoyment influence perceptions of usefulness?

**RQ3:** To what extent do perceptions of usefulness affect users’ overall evaluation of a system?

**RQ4:** What criteria do users identify when defining or describing a useful system?

### 4.2 Participants

A total of 36 participants were recruited for this study. Participants ranged in age from 18 years old to 46 years old, with a median age of 21. All participants were sophomores or above. A majority of the participants (31; 86%) were male and 5
participants (14%) were female. All participants were undergraduate students enrolled at Drexel University with majors in Information Technology, Information Systems, or Software Engineering. A majority of participants (29; 81%) reported that they had not used any of the three websites prior to completing the experiment. Six participants (16%) reported that they had previously used the ipl2 website only, and one participant (3%) reported having previously used both ipl2 and RefSeek. Nearly all participants (35; 97%) had taken at least one course in HCI; 24 participants (67%) had taken at least two HCI courses. The iKnow scores of the participants ranged from 52 to 70 (the maximum possible score) with an average score of 62.81. Unfortunately there is no data reporting iKnow scores for a general user population (Potosky, personal communication). However, other studies using the iKnow measure reported an average score of 54.3 (Potosky, 2007) which suggests that members of the study population can be considered to have “average” or “above average” knowledge of the Internet.

4.3 Marginal Model for USEFULNESS

This section will present the results of creating a marginal model with USEFULNESS as the dependent variable and all other variables as predictor variables. First, descriptive statistics for USEFULNESS are presented. Next, results from the three model creation steps described in Section 3.7.3– the initial model, covariance structure selection, and reducing the model – are explained and presented. The final model is then described and interpreted.

4.3.1 Descriptive Statistics

There was an overall mean USEFULNESS rating of 4.93 on a 7-point scale (S.D.
= 1.72), but ratings of USEFULNESS varied greatly between SYSTEMS and CONTEXTS. The descriptive statistics for USEFULNESS are presented in Table 21.

Table 21. Descriptive statistics for USEFULNESS.

<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>ipl2 Mean</th>
<th>S.D.</th>
<th>RefSeek Mean</th>
<th>S.D.</th>
<th>Awesome Lib. Mean</th>
<th>S.D.</th>
<th>Overall Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate Change</td>
<td>5.11</td>
<td>1.66</td>
<td>5.98</td>
<td>0.86</td>
<td>5.29</td>
<td>1.64</td>
<td>5.46</td>
<td>1.45</td>
</tr>
<tr>
<td>First Computer</td>
<td>3.61</td>
<td>1.99</td>
<td>5.34</td>
<td>1.66</td>
<td>3.68</td>
<td>1.95</td>
<td>4.21</td>
<td>1.99</td>
</tr>
<tr>
<td>Control</td>
<td>5.33</td>
<td>1.05</td>
<td>6.03</td>
<td>0.75</td>
<td>4.02</td>
<td>1.71</td>
<td>5.13</td>
<td>1.47</td>
</tr>
<tr>
<td>Overall</td>
<td>4.68</td>
<td>1.75</td>
<td>5.79</td>
<td>1.17</td>
<td>4.33</td>
<td>1.86</td>
<td>4.93</td>
<td>1.72</td>
</tr>
</tbody>
</table>

From the table above, the mean rating of USEFULNESS was lower for all three systems under the First Computer scenario. Additionally, RefSeek consistently received the highest USEFULNESS rating between the three systems used in the study. These differences are shown graphically in Figure 10.
4.3.2 Initial Model

The next step of the analysis was to create a marginal model with USEFULNESS as the dependent variable. The three-step top-down approach (described in Section 3.7.2.2) was used. The initial model included an intercept term and fixed effects for the main factors, CONTEXT and SYSTEM, and for four covariates: iKNOW, USABILITY, AESTHETICS, and ENJOYMENT. GOODNESS was not included since it is considered an overall evaluation of the system. Since it is unknown whether the effects of the covariates vary under different conditions, the model also includes interaction effects between the four covariates and CONTEXT and SYSTEM. Finally, a repeated measures effect is also included by specifying an unstructured covariance matrix based on the
ORDER of the repeated measurements. A summary of the variables included in the initial marginal model is presented in Table 22.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Fixed Effects</th>
<th>Repeated Measures Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>USEFULNESS</td>
<td>Intercept</td>
<td>ORDER (Unstructured)</td>
</tr>
<tr>
<td></td>
<td>CONTEXT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SYSTEM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>iKNOW</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USABILITY</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AESTHETICS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENJOYMENT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CONTEXT * SYSTEM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CONTEXT * iKNOW</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CONTEXT * USABILITY</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CONTEXT * AESTHETICS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CONTEXT * ENJOYMENT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SYSTEM * iKNOW</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SYSTEM * USABILITY</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SYSTEM * AESTHETICS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SYSTEM * ENJOYMENT</td>
<td></td>
</tr>
</tbody>
</table>

### 4.3.3 Selecting Covariance Structure for Residuals

The next step in the model-building process is selecting an appropriate structure for the residuals (repeated measures). The initial model was created using an unstructured covariance structure, meaning that the model required estimating the variances of each measurement (three parameters) and the covariance between the three measurements (three parameters). Six other covariance structures were tested using a Likelihood Ratio Test (LRT). The results of the LRTs are summarized in Table 23.
Table 23. Likelihood ratio tests of covariance structures for USEFULNESS.

<table>
<thead>
<tr>
<th>Structure</th>
<th>-2 Log Likelihood</th>
<th>$\chi^2$</th>
<th>Par.</th>
<th>p-value</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured</td>
<td>277.154</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>303.371</td>
</tr>
<tr>
<td>First-Order Autoregressive (AR(1))</td>
<td>279.902</td>
<td>2.748</td>
<td>2</td>
<td>0.601</td>
<td>288.641</td>
</tr>
<tr>
<td>AR(1)-Heterogenous</td>
<td>278.169</td>
<td>1.015</td>
<td>4</td>
<td>0.602</td>
<td>295.647</td>
</tr>
<tr>
<td>Compound Symmetry (CS)</td>
<td>279.771</td>
<td>2.617</td>
<td>2</td>
<td>0.624</td>
<td>288.510</td>
</tr>
<tr>
<td>CS-Heterogeneous</td>
<td>277.645</td>
<td>0.491</td>
<td>4</td>
<td>0.782</td>
<td>295.123</td>
</tr>
<tr>
<td>Toeplitz (T)</td>
<td>279.348</td>
<td>2.194</td>
<td>3</td>
<td>0.533</td>
<td>292.457</td>
</tr>
<tr>
<td>T-Heterogeneous</td>
<td>277.194</td>
<td>0.040</td>
<td>5</td>
<td>0.841</td>
<td>299.041</td>
</tr>
</tbody>
</table>

The LRT produced non-significant p-values (p > .05) for all alternative covariance structures, indicating that each structure provides similar explanatory power to the unstructured covariance. When selecting between similar covariance structures, it is recommended to adopt the most parsimonious structure (i.e., fewest parameters) with the lowest Bayesian Information Criterion (BIC). In the current model, the compound symmetry (CS) covariance structure is considered the best fit for the data and is therefore selected as the preferred covariance structure. A CS covariance structure contains constant variance and constant covariance. That is, the three repeated measurements are all assumed to have equal variance and are equally correlated with each other.

4.3.4 Reducing the Model

At this stage, the initial model was created with a compound symmetry structure for the residual covariance and non-significant fixed effects were removed from the model, one at a time, until only significant effects remained. The initial model contained 16 fixed effects (see Table 22). Reducing the model produced a final model with just four fixed effects (see Table 24).
Table 24. Final marginal model for USEFULNESS.

<table>
<thead>
<tr>
<th>Source</th>
<th>Numerator df</th>
<th>Denominator df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>USABILITY</td>
<td>1</td>
<td>87.119</td>
<td>283.391</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>AESTHETICS</td>
<td>1</td>
<td>82.431</td>
<td>29.856</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>CONTEXT * AESTHETICS</td>
<td>2</td>
<td>97.981</td>
<td>6.961</td>
<td>0.001</td>
</tr>
<tr>
<td>CONTEXT * iKNOW</td>
<td>3</td>
<td>98.636</td>
<td>5.513</td>
<td>0.002</td>
</tr>
</tbody>
</table>

4.3.5 Final Model

To assess exactly how each fixed effect influenced USEFULNESS, the estimated regression coefficients were examined (see Table 25).

Table 25. Coefficient estimates for the final marginal model for USEFULNESS.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>USABILITY</td>
<td>1.386</td>
<td>0.082</td>
<td>87.119</td>
<td>16.834</td>
<td>&lt; 0.001 **</td>
</tr>
<tr>
<td>AESTHETICS</td>
<td>-0.633</td>
<td>0.113</td>
<td>83.073</td>
<td>-5.613</td>
<td>&lt; 0.001 **</td>
</tr>
<tr>
<td>AESTHETICS * Climate Change</td>
<td>0.472</td>
<td>0.131</td>
<td>95.723</td>
<td>3.600</td>
<td>0.001 **</td>
</tr>
<tr>
<td>AESTHETICS * First Computer</td>
<td>0.100</td>
<td>0.119</td>
<td>99.507</td>
<td>0.836</td>
<td>0.405</td>
</tr>
<tr>
<td>AESTHETICS * Control</td>
<td>0.000</td>
<td>0.000</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>iKNOW * Climate Change</td>
<td>-0.025</td>
<td>0.008</td>
<td>98.081</td>
<td>-2.986</td>
<td>0.004 **</td>
</tr>
<tr>
<td>iKNOW * First Computer</td>
<td>0.002</td>
<td>0.006</td>
<td>98.632</td>
<td>0.370</td>
<td>0.712</td>
</tr>
<tr>
<td>iKNOW * Control</td>
<td>0.015</td>
<td>0.006100.288</td>
<td>2.342</td>
<td>0.021 *</td>
<td></td>
</tr>
</tbody>
</table>

According to the regression coefficients, CONTEXT had an indirect relationship with USEFULNESS. Specifically, higher ratings of AESTHETICS were associated with lower ratings of USEFULNESS under all three scenarios, but the effect was not as severe under the Climate Change scenario. Additionally, higher iKNOW scores were associated with lower ratings of USEFULNESS under the Climate Change scenario but higher
ratings of usefulness under the Control scenario; there appeared to be no effect of iKNOW under the First Computer scenario (likely because nearly half of all participants provided lower ratings of USEFULNESS under this scenario). Finally, higher ratings of USABILITY were associated with higher ratings of USEFULNESS under all three scenarios.

The model also produced estimates of the covariance parameters explaining the relationship between the three measurements of USEFULNESS provided by each subject. A compound symmetry covariance assumes requires estimation of two parameters: the variance of the three measurements and the covariance between the three measurements. In this case, the estimated variance was 0.47 and the estimated covariance was 0.07. The correlations between the three USEFULNESS measurements were then calculated by dividing the covariance estimate by the variance estimate. In this case, the covariance structure implies that the three ratings of USEFULNESS from each subject were only mildly correlated (r = 0.16). The estimated correlation matrix for ratings of USEFULNESS is presented in Table 26; the diagonal values are the estimated variances for each measurement while the off-diagonal values represent the correlations.

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>Second</th>
<th>Third</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second</strong></td>
<td>0.474</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Third</strong></td>
<td>0.156</td>
<td>0.474</td>
<td>-</td>
</tr>
</tbody>
</table>

| **Third** | 0.156 | 0.156  | 0.474  |

4.3.6 Summary

The final marginal model for USEFULNESS is presented visually in Figure 11. The solid lines connecting Usability and Aesthetics with Usefulness represent the
significant main effects of those two variables; the small “C” on the line from Aesthetics represents the significant Aesthetics-Context interaction. The dotted line with a small “C” connecting Internet Knowledge with Usefulness indicates the significant interaction between Internet Knowledge and Context.

![Diagram of USEFULNESS model]

**Figure 11.** Visual representation of the USEFULNESS marginal model.

In summary, ratings of USEFULNESS were significantly lower under the First Computer scenario, but the effect was smaller for participants with higher levels of Internet knowledge (iKNOW). Further, participants who rated a website as highly usable were also likely to consider the website highly useful. Meanwhile, participants who perceived a website as highly attractive were likely to view the website as less useful, although the effect was smaller under the Climate Change scenario. There was no significant relationship between perceptions of usefulness and the system or perceptions of enjoyment.
4.4 Marginal Model for GOODNESS

This section will present the results of creating a marginal model with GOODNESS as the dependent variable and all other variables as predictor variables. First, descriptive statistics for GOODNESS are presented. Next, results from the three model creation steps described in Section 3.7.2.2 – the initial model, covariance structure selection, and reducing the model – are explained and presented. The final model is then described and interpreted.

4.4.1 Descriptive Statistics

The mean GOODNESS ratings are presented in Table 27.

<table>
<thead>
<tr>
<th>CONTEXT/SYSTEM</th>
<th>ipl2 Mean</th>
<th>S.D.</th>
<th>RefSeek Mean</th>
<th>S.D.</th>
<th>Awesome Lib. Mean</th>
<th>S.D.</th>
<th>Overall Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate Change</td>
<td>5.33</td>
<td>1.72</td>
<td>6.17</td>
<td>0.94</td>
<td>4.58</td>
<td>1.98</td>
<td>5.36</td>
<td>1.69</td>
</tr>
<tr>
<td>First Computer</td>
<td>4.17</td>
<td>1.64</td>
<td>5.58</td>
<td>1.44</td>
<td>3.58</td>
<td>1.56</td>
<td>4.44</td>
<td>1.73</td>
</tr>
<tr>
<td>Control</td>
<td>5.33</td>
<td>1.30</td>
<td>5.75</td>
<td>1.06</td>
<td>3.83</td>
<td>1.64</td>
<td>4.97</td>
<td>1.56</td>
</tr>
<tr>
<td>Overall</td>
<td>4.94</td>
<td>1.62</td>
<td>5.83</td>
<td>1.16</td>
<td>4.00</td>
<td>1.74</td>
<td>4.93</td>
<td>1.69</td>
</tr>
</tbody>
</table>

There was an overall mean GOODNESS rating of 4.93 (SD = 1.69). The highest mean GOODNESS rating was for RefSeek (M = 5.83, S.D. = 1.19), followed by ipl2 (M = 4.94, S.D. = 1.62) and Awesome Library (M = 4.00, S.D. = 1.74). Mean ratings of GOODNESS across all three systems were highest under the Climate Change scenario and lowest under the First Computer scenario (see Figure 12).
4.4.2 Initial Model

The initial model for GOODNESS included an intercept term and fixed effects for the main factors, CONTEXT and SYSTEM, and for five covariates: iKNOW, USEFULNESS, USABILITY, AESTHETICS, and ENJOYMENT. The model also included interaction effects between the five covariates and CONTEXT and SYSTEM. Finally, the model included a repeated measures effect with an unstructured covariance matrix. A summary of the variables included in the initial marginal model is presented in Table 28.
Table 28. Summary of variables in initial marginal model for GOODNESS.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Fixed Effects</th>
<th>Repeated Measures Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENJOYMENT</td>
<td>Intercept  CONTEXT  SYSTEM  iKNOW  USABILITY  USEFULNESS  AESTHETICS  ENJOYMENT  CONTEXT * SYSTEM  CONTEXT * iKNOW  CONTEXT * USABILITY  CONTEXT * USEFULNESS  CONTEXT * AESTHETICS  CONTEXT * GOODNESS  SYSTEM * iKNOW  SYSTEM * USABILITY  SYSTEM * USEFULNESS  SYSTEM * AESTHETICS  SYSTEM * GOODNESS</td>
<td>ORDER (Unstructured)</td>
</tr>
</tbody>
</table>

4.4.3 Selecting Covariance Structure for Residuals

Six other covariance structures were tested using a Likelihood Ratio Test (LRT).

The results of the LRTs are summarized in Table 29.
Table 29. Likelihood ratio tests of covariance structures for GOODNESS.

<table>
<thead>
<tr>
<th>Structure</th>
<th>-2 Log Likelihood</th>
<th>$\chi^2$</th>
<th>Par.</th>
<th>p-value</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured</td>
<td>240.362</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>266.187</td>
</tr>
<tr>
<td>First-Order Autoregressive (AR(1))</td>
<td>251.462</td>
<td>11.100</td>
<td>2</td>
<td>0.025</td>
<td>260.070</td>
</tr>
<tr>
<td>AR(1)-Heterogenous</td>
<td>241.856</td>
<td>1.494</td>
<td>4</td>
<td>0.474</td>
<td>259.072</td>
</tr>
<tr>
<td>Compound Symmetry (CS)</td>
<td>251.655</td>
<td>11.293</td>
<td>2</td>
<td>0.023</td>
<td>260.263</td>
</tr>
<tr>
<td>CS-Heterogeneous</td>
<td>241.871</td>
<td>1.509</td>
<td>4</td>
<td>0.470</td>
<td>259.871</td>
</tr>
<tr>
<td>Toeplitz (T)</td>
<td>250.378</td>
<td>10.016</td>
<td>3</td>
<td>0.018</td>
<td>263.290</td>
</tr>
<tr>
<td>T-Heterogeneous</td>
<td>241.743</td>
<td>1.381</td>
<td>5</td>
<td>0.240</td>
<td>263.263</td>
</tr>
</tbody>
</table>

The LRT produced significant p-values for the AR(1), Compound Symmetry, and Toeplitz covariance structures, each of which assumes homogeneous variances between the three measurements. Models with each of those three structures produced a higher -2 Log Likelihood than the model with an unstructured covariance, suggesting that the three measurements of GOODNESS did not come from homogeneous distributions. The most parsimonious covariance structure (i.e., fewest parameters) with the lowest Bayesian Information Criterion (BIC) was the heterogeneous AR(1) covariance structure. This implies that the three repeated measurements of GOODNESS were assumed to have unequal variances and that adjacent measurements (e.g., the first and second, second and third) were more highly correlated than non-adjacent measurements (e.g., the first and third).

4.4.4 Reducing the Model

At this stage, the initial model was created with a heterogeneous AR(1) covariance structure and non-significant fixed effects were removed from the model, one at a time, until only significant effects remained. The initial model contained 19 fixed
effects (see Table 28). Reducing the model resulted in a final model with just nine fixed effects: SYSTEM, USEFULNESS, USABILITY, CONTEXT*USEFULNESS, CONTEXT*AESTHETICS, CONTEXT*ENJOYMENT, SYSTEM*iKNOW, SYSTEM*USABILITY, and SYSTEM*AESTHETICS. The final model is summarized in Table 30.

Table 30. Final marginal model for GOODNESS.

<table>
<thead>
<tr>
<th>Source</th>
<th>Numerator df</th>
<th>Denominator df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSTEM</td>
<td>3</td>
<td>64.423</td>
<td>4.986</td>
<td>0.004</td>
</tr>
<tr>
<td>USEFULNESS</td>
<td>1</td>
<td>75.310</td>
<td>38.452</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>USABILITY</td>
<td>1</td>
<td>75.435</td>
<td>10.059</td>
<td>0.002</td>
</tr>
<tr>
<td>CONTEXT * USEFULNESS</td>
<td>2</td>
<td>73.396</td>
<td>11.397</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>CONTEXT * AESTHETICS</td>
<td>2</td>
<td>72.120</td>
<td>12.169</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>CONTEXT * ENJOYMENT</td>
<td>3</td>
<td>71.686</td>
<td>6.350</td>
<td>0.001</td>
</tr>
<tr>
<td>SYSTEM * iKNOW</td>
<td>3</td>
<td>63.378</td>
<td>4.463</td>
<td>0.007</td>
</tr>
<tr>
<td>SYSTEM * USABILITY</td>
<td>2</td>
<td>74.321</td>
<td>12.015</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>SYSTEM * AESTHETICS</td>
<td>2</td>
<td>79.694</td>
<td>10.339</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

4.4.5 Final Model

The estimated regression coefficients are presented in Table 31.
Table 31. Coefficient estimates for the final marginal model for GOODNESS.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSTEM: ipl2</td>
<td>-0.626</td>
<td>1.125</td>
<td>63.940</td>
<td>-0.557</td>
<td>0.580</td>
</tr>
<tr>
<td>SYSTEM: RefSeek</td>
<td>3.590</td>
<td>1.403</td>
<td>72.986</td>
<td>2.559</td>
<td>0.013 *</td>
</tr>
<tr>
<td>SYSTEM: Awesome Library</td>
<td>-3.233</td>
<td>1.270</td>
<td>78.701</td>
<td>-2.544</td>
<td>0.013 *</td>
</tr>
<tr>
<td>USEFULNESS</td>
<td>0.574</td>
<td>0.105</td>
<td>69.474</td>
<td>5.464</td>
<td>0.000 **</td>
</tr>
<tr>
<td>USABILITY</td>
<td>-0.034</td>
<td>0.143</td>
<td>82.027</td>
<td>-0.241</td>
<td>0.810</td>
</tr>
<tr>
<td>USEFULNESS * Climate Change</td>
<td>0.146</td>
<td>0.141</td>
<td>75.811</td>
<td>1.036</td>
<td>0.303</td>
</tr>
<tr>
<td>USEFULNESS * First Computer</td>
<td>-0.408</td>
<td>0.111</td>
<td>59.484</td>
<td>-3.667</td>
<td>0.001 **</td>
</tr>
<tr>
<td>USEFULNESS * Control</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>AESTHETICS * Climate Change</td>
<td>0.630</td>
<td>0.137</td>
<td>74.884</td>
<td>4.600</td>
<td>&lt; 0.001 **</td>
</tr>
<tr>
<td>AESTHETICS * First Computer</td>
<td>0.855</td>
<td>0.150</td>
<td>79.795</td>
<td>5.699</td>
<td>&lt; 0.001 **</td>
</tr>
<tr>
<td>AESTHETICS * Control</td>
<td>0.120</td>
<td>0.175</td>
<td>62.919</td>
<td>0.684</td>
<td>0.497</td>
</tr>
<tr>
<td>ENJOYMENT * Climate Change</td>
<td>-0.379</td>
<td>0.123</td>
<td>74.311</td>
<td>-3.071</td>
<td>0.003 **</td>
</tr>
<tr>
<td>ENJOYMENT * First Computer</td>
<td>0.131</td>
<td>0.146</td>
<td>80.491</td>
<td>0.898</td>
<td>0.372</td>
</tr>
<tr>
<td>ENJOYMENT * Control</td>
<td>0.384</td>
<td>0.135</td>
<td>60.550</td>
<td>2.850</td>
<td>0.006 **</td>
</tr>
<tr>
<td>iKNOW * ipl2</td>
<td>0.000</td>
<td>0.019</td>
<td>68.793</td>
<td>0.015</td>
<td>0.988</td>
</tr>
<tr>
<td>iKNOW * RefSeek</td>
<td>-0.043</td>
<td>0.019</td>
<td>66.873</td>
<td>-2.266</td>
<td>0.027 *</td>
</tr>
<tr>
<td>iKNOW * Awesome Library</td>
<td>0.055</td>
<td>0.021</td>
<td>74.042</td>
<td>2.589</td>
<td>0.012 *</td>
</tr>
<tr>
<td>USABILITY * ipl2</td>
<td>0.803</td>
<td>0.168</td>
<td>82.620</td>
<td>4.783</td>
<td>&lt; 0.001 **</td>
</tr>
<tr>
<td>USABILITY * RefSeek</td>
<td>0.551</td>
<td>0.196</td>
<td>58.000</td>
<td>2.804</td>
<td>0.007 **</td>
</tr>
<tr>
<td>USABILITY * Awesome Library</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>AESTHETICS * ipl2</td>
<td>-0.736</td>
<td>0.180</td>
<td>86.206</td>
<td>-4.085</td>
<td>&lt; 0.001 **</td>
</tr>
<tr>
<td>AESTHETICS * RefSeek</td>
<td>-0.750</td>
<td>0.227</td>
<td>72.963</td>
<td>-3.309</td>
<td>0.001 **</td>
</tr>
<tr>
<td>AESTHETICS * Awesome Library</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

The regression coefficients indicated that USEFULNESS was the only variable to have a consistent relationship on GOODNESS; higher ratings of USEFULNESS were associated with higher ratings of GOODNESS under all three scenarios, although the effect was much weaker in the First Computer scenario. SYSTEM also had a strong effect on GOODNESS, with RefSeek receiving significantly higher GOODNESS ratings and Awesome Library receiving significantly lower GOODNESS ratings than ipl2. However, participants with higher iKNOW scores provided lower GOODNESS ratings to
RefSeek but higher GOODNESS ratings of Awesome Library. Additionally, higher ratings of USABILITY were associated with higher ratings of GOODNESS for ipl2 and RefSeek but appeared to have no effect on perceptions of Awesome Library.

According to the estimated parameters of the heterogeneous AR(1) covariance matrix, the three measurements had unequal variances of 0.32, 0.57, and 0.24 and a covariance offset of 0.14. For an AR(1) covariance structure, the correlations between the measurements can be calculated by multiplying the variance estimates by the covariance offset for adjacent measurements and by the squared covariance offset for non-adjacent measurements. This covariance structure implies that the three ratings of GOODNESS from each subject had slightly different distributions and that adjacent measurements had a very mild positive correlation (r = 0.06) while non-adjacent measurements were practically uncorrelated (r < 0.01). The estimated correlation matrix for ratings of GOODNESS is presented in Table 32; the diagonal values are the estimated variances for each measurement while the off-diagonal values are the correlations.

Table 32. Estimated correlation matrix for ratings of GOODNESS.

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>Second</th>
<th>Third</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>0.315</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Second</td>
<td>0.058</td>
<td>0.573</td>
<td>-</td>
</tr>
<tr>
<td>Third</td>
<td>0.005</td>
<td>0.051</td>
<td>0.238</td>
</tr>
</tbody>
</table>

4.4.6 Summary

The final marginal model for USEFULNESS is presented visually in Figure 13. The solid lines connecting Usefulness and System with Goodness represent the significant main effects of those two variables; the small “C” on the line from Usefulness
represents the significant Usefulness-Context interaction. The dotted line with a small “C” connecting Enjoyment with Goodness indicates the significant interaction between Enjoyment and Context. The dotted lines with a small “S” connecting Usability and Internet Knowledge with Goodness represent the significant interactions between those two variables and System. The dotted line with the small “C” and “S” connecting Aesthetics with Goodness represents the significant interactions between Aesthetics and Context and System.

![Diagram](image)

**Figure 13. Visual representation of the GOODNESS marginal model.**

In summary, a users’ overall evaluation of a system (e.g., goodness) was highly dependent on a number of factors. However, participants who perceived a website as useful were also likely to perceive the website as good. This effect was evident across all systems and contexts, although the effect was not as strong under the First Computer scenario. Conversely, the effects of usability, aesthetics, and enjoyment were dependent on the system being used and/or the context in which it was used.
4.5 Qualitative Analysis of User Responses

This section presents the results of the qualitative analysis of participants’ answers to the open-ended question, “What makes a website useful to you? What criteria do you look for?” First, data processing and preparation is explained and the data analysis method is described. Next, the four themes representing the major dimensions of usefulness and their associated codes are discussed.

4.5.1 Data Preparation and Analysis

Data from the 36 participants was split into 159 data points, where each data point represented a distinct thought or idea discussed by a participant. Overall, each participant described between 1 and 8 thoughts or concepts, with a median of 4.5. Using the constant comparative method, the data were then iteratively grouped and re-grouped until a coherent coding scheme emerged to address the research question. The coding scheme consisted of 14 distinct codes representing different concepts and one code was ascribed to each of the 159 data points. The 14 codes were then analyzed, described and grouped into four themes representing the following dimensions of usefulness: appropriateness for context, simplicity and ease of use, visual attractiveness, and pleasurable interaction. Table 33 lists the four themes and their corresponding codes.

In terms of themes identified, a majority of participants discussed more than one theme (n = 31, 86.1%), implying that the construct of usefulness consists of several different dimensions. However, only a small number (n = 5, 13.9%) discussed all four themes, which implies that all four dimensions of usefulness may not be universally valued or appreciated (at least for this particular user group). Each of the four themes will be discussed in subsequent sections.
Table 33. Overview of codes and themes from participant definitions of usefulness.

<table>
<thead>
<tr>
<th>Theme / Codes</th>
<th>No. Subjects*</th>
<th>Description of Code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appropriateness for Context (n = 28; 77.8%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suitable for Purpose/Goal</td>
<td>17 (47.2%)</td>
<td>The website provides access to desired information and helps users achieve their goal.</td>
</tr>
<tr>
<td>Right Functionality</td>
<td>13 (36.1%)</td>
<td>The website provides options for filtering, sorting, ranking, and other features to aid in search and retrieval.</td>
</tr>
<tr>
<td>Appropriate Content</td>
<td>10 (27.8%)</td>
<td>The website provides access to relevant, reliable, and accurate information.</td>
</tr>
<tr>
<td><strong>Simplicity and Ease of Use (n = 27; 75.0%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy to Use/Navigate</td>
<td>16 (44.4%)</td>
<td>The website is straightforward, logical, and efficient in terms of browsing and navigation.</td>
</tr>
<tr>
<td>Speed/Efficiency in Use</td>
<td>11 (30.6%)</td>
<td>The website provides fast and efficient access to information.</td>
</tr>
<tr>
<td>Organized/Uncluttered</td>
<td>10 (27.8%)</td>
<td>The website organizes content and information logically.</td>
</tr>
<tr>
<td>Streamlined/Simple Design</td>
<td>8 (22.2%)</td>
<td>The website is clean and simple and provides a set of visual instructions to users.</td>
</tr>
<tr>
<td><strong>Visual Attractiveness (n = 17; 47.2%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pleasing to the Eye</td>
<td>10 (27.8%)</td>
<td>The website features an appealing color scheme, appropriate graphics and readable fonts.</td>
</tr>
<tr>
<td>Craftsmanship</td>
<td>6 (16.7%)</td>
<td>The website looks and feels well-crafted and professionally designed.</td>
</tr>
<tr>
<td>General Attractiveness</td>
<td>5 (13.9%)</td>
<td>The website is has visually attractive user interface.</td>
</tr>
<tr>
<td><strong>Pleasurable Interaction (n = 17; 47.2%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>6 (16.7%)</td>
<td>The website is similar to existing websites and/or has strong brand/name recognition.</td>
</tr>
<tr>
<td>“It” Factor</td>
<td>6 (16.7%)</td>
<td>The website has a certain feel that grabs users’ attention.</td>
</tr>
<tr>
<td>Irritation-Free</td>
<td>5 (13.9%)</td>
<td>The website is free of obtrusive ads/pop-ups and is generally not frustrating to use.</td>
</tr>
<tr>
<td>Customizability</td>
<td>2 (5.6%)</td>
<td>The website has customization options that provide flexibility for users.</td>
</tr>
</tbody>
</table>

* Percentages are out of 36, the total number of participants recruited for the study. Since participants could address multiple themes in their responses, the percentages do not add up to 100%.

4.5.2 Appropriateness for Context

The most common theme that emerged from the data (n = 28, 77.8%) was that participants described a useful website as one that is appropriate for the context in which
it is being used. This theme consisted of three codes: suitable for purpose/goal, right functionality, and appropriate content. The suitable for purpose/goal code was applied to all instances in which participants indicated that a useful system should fit their specific context or purpose, specifically by helping them achieve their goal of accessing needed information. For example, one participant explained that the purpose of the website must match his/her goals:

“I guess I go to a website for a specific reason, and hopefully that function meets that specific reason, like I’m not going to a website for one of the little side things that they’re offering. I would hope that the main thing they’re offering is exactly what I’m looking for.”
[Participant 23]

Similar observations were made by other participants, with one saying “I have to ask myself why I’m there to begin with and if it satisfies that purpose it is ultimately useful” [Participant 05] and another declaring “I would want a website that could get me to what I wanted or what I needed” [Participant 13]. One participant used the example of a shopping website to express the importance of context:

“[For a shopping website] the more information they can give me to help solidify my purchase or my desire for that purchase, the more different ways they can show me that perspective I think makes it more useful. So I like ones that have a lot of reviews, overall comparisons and things like that.”
[Participant 11]

Other participants were more direct, stating “mainly the only reason I’m going to a website is usually for a specific purpose and if I can’t accomplish my goal then it’s not worthwhile” [Participant 34].

Participants noted that one aspect of being suitable for a specific purpose is offering features and functions that are appropriate and relevant to that purpose. The right
A new code was applied to all instances in which participants described website features or functions they would consider useful. Since the context of this study was information portal websites, many participants described specific components or functions of a useful information retrieval system. For example:

“I think variety is a good thing too for a useful website, because using ipl2 I typed in ‘global warming’ and then ‘politics of global warming’ but the same results came up, which is good because those results were good and they were considered good resources but a little more variety would have been helpful because I had to look for graphs [and] I didn’t look for graphs.”

[Participant 18]

Other specific search-related features mentioned were filtering results [Participant 03], expanding the number of results listed on each page [Participant 05], suggesting alternative search keywords [Participant 09], providing advanced search criteria [Participant 19], and sorting or ranking results [Participant 26]. However, some participants also discussed useful functionality in different contexts. For instance, one participant described the features of a useful shopping website:

“If I’m shopping, useful features I feel are the things you normally see on most major shopping websites. They kind of have it figured out as far as the kind of features that would be important, things like finding other related items that other people have chosen with this, showing views of what other people think about a particular product, any particular accessories I may need with it, and checkout, those are pretty standard.”

[Participant 09]

Similarly, some participants discussed the importance of a website’s content, with many participants focusing again on the context of information retrieval. The appropriate content code was applied to all instances in which participants described aspects of the resources or information returned by an information retrieval system. As one example:
“If I had a question I would type in the main keywords and it would bring up some topics about it. The first one [Awesome Library] I struggled with, because when I tried to type in three different options...it didn’t have anything relevant to [my search terms].”
[Participant 17]

Other participants also referred to relevant search results as a key indicator of a useful search-related website, observing, for example, “it wouldn’t be a good website without good resources too” [Participant 20] and “if I search for something, I want it to give me something close to my search” [Participant 28]. Another aspect captured by this code was the accuracy and credibility of the sources provided by the search engine:

“If it’s supposed to be a site that’s geared towards students, I wouldn’t want wiki links popping up because students are told not to use those websites. So I think the information needs to be credible, and that would make the site useful.”
[Participant 22]

Other participants had similar sentiments, with one participant noting that “I’d say what sources are actually returned… it wouldn’t be a good website without good resources” [Participant 20] and another stating that “making sure…the content reputability is appropriate for what I’m looking for” [Participant 06].

4.5.3 Simplicity and Ease of Use

The second most common theme that emerged from the data (n = 27, 75.0%) was that participants described a useful website as one that is both simple and easy to use. This theme consisted of four codes: easy to use/navigate, speed/efficiency in use, organized/uncluttered, streamlined/simple design. The easy to use/navigate code was applied to all instances in which participants explained that a useful website should be easy to learn and navigate. For example, two participants indicated that a useful website should require little effort to learn and use:
‘You don’t have to sit in front of a webpage for a long time figuring out, all right, how do I want search this thing here? The steps you’re taking yourself in your mind, you can do the same thing here on the webpage. So that’s also very important.”
[Participant 10].

“A useful website would be, so if I’m going to look for a certain topic or if I want, you know, news or something, I’d like it to be pretty up front in the sense that I don’t have to put out much effort, I don’t have to dig down a lot, I don’t have to do a lot of work to find what I’m looking for.”
[Participant 08]

Other participants discussed ease of use in terms of providing logical and efficient navigation. For example, one participant observed that “I would define a useful website where one I could navigate easily and quickly” [Participant 22] while another said that a useful website was one where s/he could “navigate it easily” [Participant 24]. Another explained that a useful website is “just straightforward, so immediately when I log on I’m able to identify what I need, navigate to it and just get it straight up” [Participant 02].

A similar and widely cited aspect of ease of use was quickness. The speed/efficiency in use code was applied to all instances in which participants described a useful website as one that allows them to quickly and efficiently access the information they need. The importance of speed was perhaps best captured by this participant:

“And I guess the other thing is just quick and easy to use. You don’t want to spend too much time on the site to get the search results, regardless of where it’s pulling from, get them as quick as possible, pull all data as quick as possible, even if you’re not doing search results. If you’re looking at stats of something, you want it all to pull as quick as possible.”
[Participant 23]

The importance of speed and quickness was also stressed by a number of other participants. For example, one participant stated that “just getting what you need as fast
as you can get it” [Participant 06] was of critical importance. Another participant likewise remarked that a useful website was one where “I don’t have to spend a lot of time looking around to get to what I’m trying to do” [Participant 24].

Some participants explained that they value efficiency because there are often multiple things competing for their attention. As one participant noted “I want what I want and then I want to move on to the next thing I’m interested in” [Participant 12]. Another stated “I’m one of those people that, I don’t like to stay on one webpage for a long time. So… I want it to be as quick as possible to get me where I want to” [Participant 10].

A related topic was the organization of the website, which refers to both its structure and its appearance (as one participant explained, “it would have to be organized [and] it would have to look organized” [Participant 11]). The organized/uncluttered code was applied to all instances in which participants discussed the overall structure and organization of a website. For example, one participant explained that a website should be organized logically in order to map to his/her thought process:

“Categorization helps, because in your mind you’re breaking down information, you know you can start with the broader picture, you’re breaking it down. If the website is designed that it’s almost doing that with you, you know, the process becomes much simpler.” [Participant 10]

Another participant also stressed the importance of a logical layout, saying that a useful website should be “arranged in a way that they [users] can follow it through and not be confused by which part to go through” [Participant 16]. Other participants discussed organization in negative terms, associating a poor website with excess clutter (e.g., lack of organization):
“And not something that just seems so cluttered with a lot of different types of information on the website, like students and teachers when you’re probably not going to need that it should probably be on the side somewhere or at the bottom where you won’t see it. It’s too much.”
[Participant 36]

Several participants made similar points, with one stating that a useful website should be “not cluttered, but very organized” [Participant 14] and another noting that “you want to be able to see what you’re looking for. You don’t want it to be too cluttered” [Participant 23].

The final aspect of this theme is related to the layout of the website. The streamlined/simple design code was applied to all instances in which participants discussed the importance of having a clean and simple interface that provides a set of “visual instructions” to users. As one participant explained, simplicity is critical:

“I think simplicity is definitely a big part of design, trying to make sure that people can find stuff and it just has to be laid out in a way that people can see it. For example if the search box is the most important thing on your website, you should make it stand out and actually jump out at you so you know where to go to first start looking. For example, on a banking website you can’t have a log-in on the very bottom where nobody can see it because that’s probably the first thing people are going to go to when they’re trying to check their account.”
[Participant 19]

Several others agreed, with one participant stating that a useful website should be “something that is really clean looking” [Participant 36] and another saying “I’m a big fan of simplistic design, like Google - no flashy banners or CSS and flash and all that” [Participant 01]. Some participants described a streamlined and simple design as one that provides a sort of visual instructions for how a website should be used; in other words, “it’s got to have the natural mapping, like when you look at something your eyes need to
be guided toward where you want the information to be” [Participant 23]. Another participant defined a useful website as one that helps to “visually sort” the information presented:

“When I say visually sorted it doesn’t have to have pictures per se of things but maybe things that are in a logical order of viewing something…Headers more at the top, with a progressively large thing going to smaller towards the bottom, and making sure that there are links to go back to the original start of the website. Those are things that I look for and that are visually appealing and useful for me.” [Participant 25]

4.5.4 Visual Attractiveness

The third theme that emerged from the data (n = 17, 47.2%) was that participants described a useful website as one that is visually attractive or aesthetically pleasing. This theme consisted of three codes: pleasing to the eye, craftsmanship, and general attractiveness. The *pleasing to the eye* code was applied to all instances in which participants referenced a website’s color scheme, graphics, font, or otherwise discussed a website in terms of its visual elements. For example, one participant stated that a useful website should be “easy to read and easy on the eyes” [Participant 21]. Another similarly stated that “you want a design that’s attractive, you don’t want it to be hard on the eyes” [Participant 23]. Yet another remarked that “it just needs to be appealing to the eye” [Participant 13]. Other participants mentioned specific visual elements, such as “good colors and good graphics” [Participant 33], “being able to read clearly, not having a small font” [Participant 35], and “properly color coded buttons” [Participant 32].

Aside from its visual elements, participants also discussed website aesthetics in terms of having a professional appearance. The *craftsmanship* code was applied to all instances in which participants discussed aspects of how well-crafted and professionally
designed the website seemed to be. Here’s how one participant explained the importance of a well-crafted website:

“[It] kind of stems from trust and confidence in the website, which a lot of visuals play into that. If it looks well put together I’m assuming that professionals put it together and have curated the content, so that lets me know that the source is credible or that the people that they’re citing are credible. I don’t know, it’s kind of a weird system of trust.” [Participant 04]

Other participants made similar points, with one stating that a clean, professional layout makes them “feel like [designers] probably put more effort into putting out the right sources and functions and everything” [Participant 20]. Another participant made a related point in discussing the aspects of “modern” design, saying “I’m not sure what they call it, the Apple style of buttons like in the iPhone navigation…I think it makes the website look really professional” [Participant 26].

Finally, the general attractiveness code was applied to all instances in which participants made vague or general comments about a website’s attractiveness. Some examples were “how great visually it is” [Participant 12], “how good it looks” [Participant 13], “it looks nice” [Participant 14], and “it should have a nice UI” [Participant 26].

4.5.5 Pleasurable Interaction

The fourth theme that emerged from the data (n = 17, 47.2%) was that participants described a useful website as one that is both pleasurable and enjoyable to interact with. This theme consisted of four codes: familiarity, “it” factor, irritation-free, and customizability. The first aspect of providing pleasurable interactions is taking advantage of users’ experiences to give them a sense of comfort and familiarity. The familiarity code was applied to all instances in which participants discussed aspects of a website that
were consistent with existing websites or discussed the name/brand recognition of the website itself. Because the context of the study was search and information retrieval, several participants drew comparisons to popular search engines, specifically Google. As one participant explained, the RefSeek website was considered useful because “it looked like Google” [Participant 31] while another stated that a useful website is “something that looks like Google, like plain white with just a search box and search button” [Participant 36]. Another aspect of familiarity relates to the corporation, organization or institution hosting the website. As one participant explained, unknown sites can only be useful if they convince users they are credible and trustworthy:

“This one here [ipl2]...I saw that it’s partnered with all these other schools. All the resources on here are legitimate, it has the ‘connect with it’, you can tell right away what it is, it’s kind of a depository [sic] for a bunch of scholarly resources, featured articles, help, site map, none of the other sites had a site map, that’s big. The privacy policy, contact, and then it links to all the different schools here, so that alone helps the website immensely.” [Participant 03]

Another participant explained that his/her perception of a website’s usefulness is based on “some portion of it is its name, what it’s actually known for, its reputation” [Participant 19]. A final participant took this idea even further, suggesting that a site’s reputation is important even if that reputation is somewhat negative:

“If I wanted to learn a lot about information I’d probably go to Wikipedia... I don’t know if that information is correct, but I’m learning something – it might be incorrect, but it’s there and I know it’s there and it’s something I can rely on.” [Participant 06]

A second aspect of pleasurable interactions referred to the overall “feel” of the website. The “it” factor code was applied to all instances in which participants described
a useful website in terms of what Alexander (1979) referred to as the “quality without a name.” As one participant explained:

“It’s definitely when you first visit the webpage, how…what kind of feel that gives you, if it feels overwhelming, I might not use it, if it feels like just a search, and Google does that best, it’s just a search – if you want other options they’re there. So it’s all about that kind of feel when you get there.” [Participant 19]

Another participant echoed these sentiments, saying “just everything that has this…how the website flows” [Participant 15]. One participant discussed “feel” by describing what a website should not do in terms speed, visual design, organization, and purpose:

“If the website is slow, or if there’s too many images, or too many things cluttering it and there’s no focal point, and I can’t find what I need easily, then I kind of just lose my attention span and I’ll try another one.” [Participant 13]

Other participants were more specific, stating that a useful website should be “pleasant to use” [Participant 08] or that it should be interactive “like how Google search engine sometimes with the logo they do cool things like the guitar… [it] helps attract people to your website, it gets spread around word of mouth” [Participant 10].

A third aspect of providing pleasurable interaction is making sure users are not frustrated or annoyed when using the website. The irritation-free code was applied to all instances in which participants discussed undesirable features of a website (e.g., features that restrict its usefulness). In these instances, the most prevalent issue was related to whether or not the website places barriers between the user and the information they want to access. As one participant explained, pop-up ads or full-screen advertisements are distracting and frustrating, even if some of them are useful:
“A useful website: one that doesn’t have too many Google ads, one that doesn’t have pop-up ads, although those can still be useful – I still use websites that have pop-up ads, not exactly ‘pop-up’ ads in a separate window but they take over the website and black out the rest of the screen.” [Participant 11]

Other participants also made reference to pop-up ads, with one participant saying that a useful website should be “unobtrusive – I don’t want 50 pop-ups coming up, I don’t like those grayed out boxes where it blurs out the page” [Participant 32]. Another participant broadened the issue of pop-up ads to include distracting color schemes, noting “you want the colors to mesh in the sense that what you’re looking for will stand out above everything else, it won’t be hidden by pop-ups and advertisements and other irrelevant information” [Participant 23].

One final aspect of pleasurable interactions is providing flexibility for users through customization options. The customizability code was applied to all instances in which participants discussed being able to adapt or tailor a website based on their personal tastes or interests. Only two participants mentioned this issue, with one stating that “the key, at least for me, is a site that allows me to specifically choose what it is that I’m looking for or be able to specify how I want the information displayed” [Participant 09] and the other saying a website is useful “if it has anything that’s customizable to my views and stuff that I’m interested in” [Participant 34].

4.5.6 Summary

This section described the results of a qualitative analysis of participants’ thoughts and opinions about the criteria for defining a useful website. Using the constant comparative method, the data were first broken into 14 codes representing distinct ideas and concepts. Next, the 14 codes were grouped into major themes representing the major
dimensions of usefulness identified by participants: appropriateness for context, simplicity and ease of use, visual attractiveness, and pleasurable interaction. The four themes and their associated codes are depicted graphically in Figure 14.

![Diagram of four dimensions of usefulness](image)

**Figure 14.** Four dimensions of usefulness identified by participants

### 4.6 Summary

This chapter described the results of a repeated measures experiment to investigate the concept of system usefulness – defined in terms of the functions it provides, the tasks or goals it supports, and the context in which it is used – and its relation to context and major components of UX (usability, aesthetics, and enjoyment). In the experiment, 36 participants interacted with three different information portal websites in three different contextual scenarios and were then asked to complete a 17-item post-
scenario survey rating each system in terms of usefulness, usability, aesthetics, enjoyment, and goodness (their overall evaluation of the system). Quantitative results showed that 1) usefulness was highly dependent on context, 2) usefulness was closely related to usability, 3) usefulness was negatively related to aesthetics, 4) usefulness was unrelated to enjoyment, and 5) usefulness was a critical element of a user’s overall system evaluation. Additionally, participants were asked to provide their own definition of what constitutes a useful website. Their responses were analyzed using the constant comparative method and four major themes emerged: appropriateness to context, simplicity and ease of use, visual attractiveness, and pleasurable interaction. The findings from the quantitative and qualitative analyses will be discussed further Chapter 5 to determine whether the results reinforce or contradict each other.
5. DISCUSSION

This chapter synthesizes the results of the quantitative and qualitative analyses presented in Chapter 4. First, the previous chapters are reviewed and summarized. Next, the findings are discussed in relation to the four research questions regarding usefulness and context, usefulness and usability, aesthetics, and enjoyment, usefulness and goodness, and user definitions of usefulness.

5.1 Introduction

A review of the literature showed that the meaning of usefulness has changed over time, from a narrow system-based focus on reliability to a broader user-centered focus on ease of learning and usability. With user experience (UX) emerging as the dominant research paradigm in Human-Computer Interaction (HCI), it is likely that the meaning of usefulness must be updated accordingly. However, despite its importance, there is little consensus about a precise definition of the term “usefulness” and studies of usefulness have been largely absent from HCI research. Drawing from the literature, a working definition of usefulness was proposed as “the extent to which a system’s functions allow users to complete a set of tasks and achieve specific goals in a particular context of use.” The literature also suggested that perceptions of usefulness may be influenced by changes in context and by the usability, aesthetics, and enjoyment of the system, but there is little empirical evidence describing these relationships. Thus, the following four research questions were addressed:

RQ1: How and to what extent are perceptions of usefulness shaped by the context—the user, task, and environment—in which the system is used?
**RQ2**: How and to what extent do perceptions of usability, aesthetics, and enjoyment influence perceptions of usefulness?

**RQ3**: To what extent do perceptions of usefulness affect users’ overall evaluation of a system?

**RQ4**: What criteria do users identify when defining or describing a useful system?

Consequently, a repeated measures experiment was designed to investigate this concept of usefulness and its relation to context and major components of UX (usability, aesthetics, and enjoyment). In the experiment, 36 participants interacted with three different information portal websites in three different contextual scenarios and were then asked to rate each system in terms of usefulness, usability, aesthetics, enjoyment, and goodness (their overall evaluation of the system). Results of the quantitative analysis showed that 1) usefulness was highly dependent on context, 2) usefulness was closely related to usability, 3) usefulness was negatively related to aesthetics, and 4) usefulness was unrelated to enjoyment. It was also found that usefulness was a critical element of a user’s overall system evaluation. Additionally, participants were asked to provide their own definition of what constitutes a useful website and four major themes emerged: appropriateness to context, simplicity and ease of use, visual attractiveness, and pleasurable interaction. Each of these findings will be discussed further in subsequent sections.

**5.2 RQ1: Usefulness and Context**

**RQ1**: How and to what extent are perceptions of usefulness shaped by the context—the user, task, and environment—in which the system is used?
The first research question addressed the relationship between usefulness and context. This research borrows from Shackel (1991) and the ISO (1997) to consider context in terms of its main elements: the user, the system, the task, the physical environment and the social/cultural environment. RQ1 was answered by determining the effects of CONTEXT on USEFULNESS in the quantitative analysis and by examining whether participants identified contextual factors as important components of usefulness in the qualitative analysis.

The results of the marginal model showed that participants’ ratings of usefulness were influenced by the interaction between context and Internet Knowledge. There are three implications for this finding: first, the effects of context are not constant and likely cannot be determined a priori without an in-depth understanding of a system’s users; second, users’ technical expertise plays an important role in determining how their perceptions of system usefulness are shaped by the context in which the system is used; and third, that individual differences have a strong influence on user perceptions of usefulness, even when those differences appear to be minor (all participants in this study had above average Internet knowledge). Furthermore, the most widely cited aspect of usefulness discussed by participants was whether the system was appropriate for their specific context in terms of 1) helping them achieve their goal, 2) providing appropriate functionality, and 3) giving them access to relevant content. Additionally, it was shown that the First Computer scenario produced the lowest mean ratings of usefulness and the highest variance in ratings of usefulness across all three systems. The First Computer scenario asked users to find a very specific piece of information (the name, inventor and invention date of the first computer) and success was determined by whether or not
participants felt they were able to find a reputable source containing that information. By design, this scenario was close-ended and provided few potential pathways for completion. Some users may have had a more difficult time completing the scenario than others since only a limited number of different searching strategies would result in successful completion of the scenario, leading to a wider range of usefulness ratings and lower ratings of usefulness overall.

These findings confirm the suspicion that the usefulness of a system cannot be separated from the context in which it is used. This result is not particularly surprising for a number of reasons. First, the relationship between usefulness and context has long been understood. In fact, this relationship was perhaps first noted by the Roman architecture critic Vitruvius who described *utilitas* as “the appropriate allocation of space in a design” or whether a building is “suitable for the purposes for which it is intended” (Hong & Kim, 2004, p. 338). However, this study is the first to demonstrate the strength of this relationship through an experimental approach.

Second, the working definition of usefulness proposed in this study included an explicit connection to contextual factors, suggesting the results of the quantitative analysis may have been tautological. However, the connection between usefulness and context was also confirmed by the qualitative analysis, which showed that 80% of participants in the study defined usefulness in terms of being suitable for their specific context. So while not providing a new result, the significant effect of context reinforces the notion that contextual factors need to be considered when studying usefulness.

Third, it is a somewhat obvious observation that some systems are more appropriate for certain contexts than other systems. For instance, the website of a sporting
goods store would obviously not be as useful as a travel website of if a person is interested in making flight and hotel reservations. But in this experiment, the two contextual scenarios (excluding the control scenario) presented two different types of information retrieval problems and all three systems provided access to similar types of information resources. Yet still, all three systems were rated as least useful under the First Computer scenario, and other variables (aesthetics, Internet knowledge) had different effects on usefulness depending on the context in which the system was used. These results suggest that the relationship between context and usefulness is more nuanced than simply whether a user can perform specific tasks in an environment. These results was supported by the qualitative analysis, which showed that participants also defined a useful website in terms of whether it provides the “right” functions and provides access to the “right” information, where the “right” is determined entirely by the user’s background, purpose and goal in visiting the website.

5.3 RQ2: Usefulness and Usability, Aesthetics and Enjoyment

*RQ2: How and to what extent do perceptions of usability, aesthetics, and enjoyment influence perceptions of usefulness?*

The second research question addressed the relationship between perceptions of usefulness, usability, aesthetics and enjoyment, all of which were measured using multi-item rating scales on the post-scenario survey. Therefore, RQ2 was answered by looking at the quantitative relationships between the four variables and by examining the qualitative data to see whether participants discussed issues related to usability, aesthetics, or enjoyment when defining a useful website. The results showed that usefulness and usability are closely linked, that usefulness has a nuanced relationship
with aesthetics, and that usefulness may only be tangentially related to enjoyment. Each of these findings will be discussed in the following sections.

5.3.1 Usefulness and Usability

According to the results of both the quantitative and qualitative analyses, the concepts of usefulness and usability are closely linked. Specifically, the results of the quantitative analysis showed that participants who perceived a website as highly usable also perceived it as highly useful. This relationship was also supported by the qualitative analysis in which the major components of usability—efficiency (speed), effectiveness (goal/purpose), and satisfaction (irritation-free/"it" factor)—all emerged as commonly cited criteria defining a useful website.

Once again, this result was not particularly surprising because of how the two terms were defined in this experiment: usefulness as a system’s fit for a specific context and usability as its effectiveness, efficiency and satisfaction within that context. However, several researchers have noted that usability issues can either enhance or inhibit usefulness by ensuring there are no barriers between the users and the functions or content they need (Long, et al., 1983; Eason, 1994; Keil, et al., 1995, Abdelnour-Nocera, et al., 2007). Furthermore, 75% of all participants referred to issues related to usability and ease of use when offering their definition of a useful website. In particular, participants defined a useful website as one that is easy to navigate, allows them to get what they need quickly, is logically organized and without excess clutter, and has a clean, streamlined design. Therefore, it is likely that the close relationship between usability and usefulness is not merely a byproduct of using these definitions but rather a confirmation that usability is still a critical factor in defining a useful website.
Since the two constructs were so similar, it is reasonable to consider whether the construct of usefulness is worth considering at all; after all, if participants perceived a usable website as a useful website, is the construct of usability sufficient for effective evaluation? To answer this question, it should be considered whether traditional usability evaluation methods would adequately address the dimensions of usefulness raised by participants in this study. To do so, the four dimensions of usefulness and their associated codes were compared with three popular methods of measuring or assessing usability: Norman’s Design Principles (Norman, 1988), Nielsen’s Usability Heuristics (1994), and Shneiderman’s usability measures (1997). This comparison is presented in Table 34.
Table 34. Comparing dimensions of usefulness with usability measures.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriate for Context</td>
<td>Suitable for Purpose/Goal Right Functionality Appropriate Content</td>
<td>Mapping</td>
<td>Match btw. system/world Help and documentation Visibility of sys. status</td>
</tr>
<tr>
<td>Simplicity and Ease of Use</td>
<td>Easy to Use/Navigate Mapping</td>
<td>Feedback Speed/Efficiency in Use Feedback Speed Efficiency of use Minimalist design</td>
<td>Error prevention Efficiency of use Minimalist design</td>
</tr>
<tr>
<td>Visual Attractiveness</td>
<td>Pleasing to the Eye Pleasing to the Eye</td>
<td>Constraints Constraints</td>
<td>Recognition v. Recall</td>
</tr>
<tr>
<td>General Attractiveness</td>
<td></td>
<td>Affordances</td>
<td></td>
</tr>
<tr>
<td>Pleasurable Interaction</td>
<td>Familiarity Consistency Consistency/standards</td>
<td>“It” Factor Consistency Error diagnostics</td>
<td>Irritation-Free Satisfaction Use control/freedom Customizability</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Not surprisingly, the table shows that a majority of measures or principles commonly associated with usability are concentrated around simplicity and ease of use.

Some measures indirectly assess whether a website provides pleasurable interactions but there are no measures or principles that assess a website’s visual attractiveness or suitability for a specific purpose. It could be argued that user testing with scenarios would adequately address whether a system is suitable for its purpose, but the success of these
methods may depend on whether or not evaluators selected the right scenarios for use in the evaluation.

It is no surprise, then, that traditional usability evaluation methods are poorly suited for capturing or investigating factors related to usefulness (Greenberg & Buxton, 2008). Dicks (2002) noted that one the major problems of usability testing is that it emphasizes ease of use at the expense of usefulness because the task-centered nature of usability testing does not provide insight into how well users are able to grasp the overall complexity of a system. He observed that:

“Our [usability] results have shown that users can complete every task assigned on a test within reasonable times. However, they often report that they do not like the overall site and would not use it again... it has often been due to larger, overall usefulness problems associated with the sites” (Dicks, 2002, p. 29).

These issues were also discussed by Nørgaard and Hornbæk (2006) who observed that problems of utility were addressed infrequently during usability testing because evaluators rarely questioned or further investigated statements made by participants about the utility of the system. When evaluators did address utility problems, it was usually through ad-hoc, informal conversations with users (Nørgaard & Hornbæk, 2006). One potential exception to this rule was mentioned briefly as part of the CUE-4 study described by Molich and Dumas (2008). In the study, one team (team G) reported using a methodology called “persona-based review,” which was described as similar to a heuristic evaluation except that experts were asked to compare the product to the needs of a pre-defined set of personas, rather than a set of principles. The team members claimed that this method defined the context of use and “is more likely to discover problems related to usefulness, not simply problems related to usability” (Molich & Dumas, 2008,
However, the number and type of problems reported by Team G were not significantly different from those reported by any other team and the authors did not report on whether Team G’s results were qualitatively different than those of any other team. Based on these observations and the results of this study, it is reasonable to conclude that usability evaluation methods alone are not well-suited for evaluating usefulness. There are two reasons for this conclusion: first, traditional methods focus on task-centered user performance rather than a system’s fit to context (i.e., whether it offers the right functions to achieve users’ goals); second, the usefulness of a system may be determined by more than just its usability and fit to context; users also define usefulness in terms visual attractiveness and the provision of pleasurable interactions.

A separate, but related, issue is how perceived usability/usefulness is related to actual usability/usefulness. In HCI research, perceived usability is generally captured by asking users whether they believed a system was easy to use, easy to learn, etc. (e.g., Hassenzahl & Monk, 2010) while actual usability is captured using traditional usability metrics like task completion time, task completion rate, and number of errors (Hornbæk & Law, 2007). While it can be argued that the metrics used to measure actual usability are less biased and more reliable, some studies have shown that perceived usability is actually more important. For instance, Frøkjær, Hertzum, and Hornbæk (2000) found that users overwhelmingly preferred a system that allowed the use of multiple interaction styles even though users were more efficient and more effective when using only a single interaction style. Similarly, Gong and Lai (2001) reported that users preferred interacting with a mixed-speech system (human and synthetic) compared to a synthetic-only system, even though users performed significantly better using the synthetic-only system.
Likewise, Hornbæk and Frøkjær (2001) observed that 95% of users preferred to read electronic documents using a combination of two interaction styles even though they spent significantly more time reading and answering questions when using it, and Corbett and Anderson (2001) discovered that even though students performed significantly better when they were given more feedback, there were no significant differences in terms of how students perceived tutoring systems with differing levels of feedback and students showed no strong preference for any of the different tutoring systems. From these examples, it seems clear that users’ perceptions of system usability are not always aligned with the actual usability of the system, and that their perceptions are often the driving force in their decision-making process. One explanation for this phenomenon was provided by Don Norman:

“I strongly believe that for most, everyday applications (products, applications, websites, and web applications), perception is more important than reality. [For most people], something that takes longer but that is perceived to be efficient is superior to something that is shorter but perceived differently.” (Norman, 2004)

While the distinction between perceived and actual usability is clear, the distinction between perceived and actual usefulness is more complicated. While researchers and practitioners in HCI know a great deal about measuring actual usability, there is little knowledge of how to measure actual usefulness. If actual usability is evaluated using usability metrics, it follows that actual usefulness should be evaluated using usefulness metrics. But what exactly are these usefulness metrics? The challenge in finding these metrics is that “actual usefulness” is determined by 1) the system being assessed and 2) the relevant contextual factors surrounding its use (including the characteristics of its users). For instance, the usefulness of an educational system might
be defined in terms of its impact on student learning while the usefulness of an e-commerce system might be defined in terms of how many customers reported high levels of satisfaction. But, as with usability, perceptions of usefulness may be more important than actual usefulness since users are only going to continue using a system if they perceive it to be useful (regardless of whether it is actually useful). Therefore, designers should not be content with evaluating systems for actual usefulness alone. Instead, they should strive to build systems that are both actually useful and perceived as useful by users because, as Norman (2004) argued, “what [people] care about is: did they get the answer or outcome they required; was it painless; was it enjoyable. Would they do it again (voluntarily, not because they had no choice)?”

5.3.2 Usefulness and Aesthetics

The results of this study showed that there is a somewhat complicated relationship between usefulness and aesthetics. The results of the quantitative analysis showed that aesthetics had a significant negative relationship with usefulness, suggesting that users may actually perceive a website as less useful if they also perceive it as attractive. This finding conflicts with the assertion that users tend to view attractive systems more favorably in other aspects (e.g., the “halo effect”: De Angeli, et al., 2006). However, although this negative effect was evident for all systems under all contextual scenarios, it was not quite as strong under the Climate Change scenario. One possible explanation for this result is that participants had little difficulty completing the Climate Change scenario because it was open-ended and had many pathways to success. Therefore, it is possible that a positive experience under the Climate Change scenario resulted in a less critical assessment of the system’s aesthetics. This suggests that the influence of aesthetics on
usefulness is not constant, and that the two concepts may exhibit a different relationship in different contexts.

By contrast, the results of the qualitative analysis showed that visual attractiveness was cited by almost 50% of participants as an important component of usefulness. Specifically, participants asserted that the attractiveness of a website was important in determining its usefulness because it made the site seem more professional and they mentioned several aspects of visual design that they felt were important, including color balance and brightness, font size and type, use of images, and a generally attractive aesthetic (e.g., using modern visual design elements like shading, shadows, and other elements that provide the illusion of three-dimensionality).

Taken together, the results of this experiment show that although ratings of aesthetics had a negative effect on ratings of usefulness, users still believed that visual attractiveness was an important element of usefulness. One explanation for these conflicting results is the “beauty dilemma” (Diefenbach & Hassenzahl, 2009). In a study of mobile phone attractiveness, Diefenbach and Hassenzahl observed that although aesthetics was a valued aspect of design, users discounted its importance when making product decisions and often did not cite it as justification for making such decisions. This phenomenon was termed the “beauty dilemma” because even though visual attractiveness often results in a positive interactive experience, it is still considered mostly as “peripheral, ambiguous and soft nature, which makes it hard to justify” (p. 1420). Similarly, in this study participants were asked to make a quantified aesthetic judgment as a series of numeric ratings on the post-scenario survey. Since beauty or attractiveness is rarely assessed in this manner, this process may have been similar to making a product
decision and could have caused some participants to subconsciously de-value aesthetics because of their desire to appear rational (Diefenbach & Hassenzahl, 2009). Meanwhile, when they were asked to verbally describe the criteria they look for in a useful website, participants likely felt more comfortable talking about the importance of aesthetics because they were able to talk about it in more ambiguous terms (e.g., “you don’t want it to be hard on the eyes” [Participant 23]). However, participants still felt the need to justify their opinion in rational terms by explaining that aesthetics was important because it either made the website easier to use or that it lent the website a sense of professionalism and credibility.

In summary, this study did confirm the existence of a relationship between usefulness and aesthetics but unfortunately the exact nature of this relationship remains unclear. The results of this study may have been because the user group chosen for this study–technically-savvy college students–are more likely to appreciate technical aspects (e.g., speed, processing power, functionality) over aesthetic aspects. However, users in this group also grew up in the “Web 2.0” environment and have become accustomed to high-quality visual design due to their familiarity and comfort with products (e.g., the Apple iPhone) and systems that emphasize visual design and attractiveness as core features. Future studies with different user groups will further investigate whether and how visual attractiveness influences usefulness and whether it is universally valued, dependent on the user group, or simply a matter of personal preference.

5.3.3 Usefulness and Enjoyment

In terms of enjoyment, the quantitative results showed that users who found systems useful did not also find them enjoyable. By contrast, providing a pleasurable
interaction was mentioned by almost half of participants as an important aspect of a
useful website, with participants mentioning that a useful website should feel familiar and
consistent, that it needs to have the right “feel” or “flow,” that it should be free of
irritating features like pop-up windows or obtrusive ads, and that it should offer
flexibility through customization options. There are three possible explanations for these
seemingly conflicting results.

First, it is possible that “enjoyment” and “pleasurable interactions” actually refer
to two different (but related) concepts. The two questions measuring enjoyment on the
post-scenario survey were “I find using the website to be enjoyable” and “I have fun
using the website.” When answering these questions, many participants explained that
they were hesitant to use words like “enjoyable” and “fun” because they associated those
terms with more leisurely activities like playing video games or spending time with
friends. In other words, participants did not “enjoy” using the websites because they
don’t typically “enjoy” using websites at all, especially when using them for educational
purposes. By contrast, the “pleasurable interaction” theme that emerged from the
qualitative analysis referred specifically to certain aspects of using websites that
participants liked and/or preferred; in fact, no participants actually used the terms “fun,”
“pleasure/pleasurable,” or “enjoy/enjoyment” when describing a useful website.
Therefore, it is likely that there is some type of hedonic/pleasurable component of
usefulness but that it is more nuanced than the traditional notion of “enjoyment” typically
associated with UX.

A second and related explanation is that the non-significant quantitative
relationship between enjoyment and usefulness was a byproduct of the experimental
design. First, it is possible that the three websites used in the study were not enjoyable to
use under any circumstances either due to their purpose as information portals or due to
poor UX design in general. Repeating the experiment with different websites that are
more interactive or engaging may yield different results. Second, it’s possible that the
contextual scenarios used in the study did not allow for an enjoyable interactive
experience due to their subject matter (computer history and climate change), purpose
(educational assignment and small friendly wager), or the mundane nature of the tasks
provided (finding resources for a PowerPoint presentation and finding the inventor of the
first computer). Once again, repeating the experiment with contextual scenarios in
different domains and with more exciting or engaging tasks may yield different results.
Finally, it’s also possible that the five minute time limit for interacting with each website
prevented participants from developing stable and accurate perceptions of enjoyment.
Additional experiments that include more prolonged exposure to the websites,
supplemented with longitudinal studies (e.g., Karapanos, et al., 2009), may provide more
insight into this concept.

A third explanation is that this study exhibited differences between what
Hassenzahl and Monk (2010) called “do-goals” (e.g., pragmatic quality) and “be-goals”
(e.g., hedonic quality). In this study, usefulness was considered only in terms of “do-
goals” in that it was measured on the post-scenario survey in terms of how well the
system allowed users to complete tasks. But pragmatic quality is only half of the equation
and is entirely separate from hedonic quality, or “a product’s potential to support pleasure
in use and ownership” (Hassenzahl & Monk, 2010, p. 236). From this perspective, the
results of this study suggest that system usefulness should not be defined solely in terms
of pragmatic quality but should also include hedonic attributes, such as its ability to provide pleasurable interactions.

5.4 RQ3: Usefulness and Goodness

**RQ3**: To what extent do perceptions of usefulness affect users’ overall evaluation of a system?

The third research question addressed the relationship between USEFULNESS and a user’s overall evaluation of a system, which in this experiment was represented by the GOODNESS variable. Therefore, RQ3 was answered by examining the results of the quantitative relationships between GOODNESS and USEFULNESS.

According to the results of the quantitative analysis, usefulness was the only variable that had a significant positive effect on ratings of goodness across all systems and contexts (although the effect was not as strong under the First Computer scenario). Essentially, websites that are perceived as useful are also perceived as good. Once again, it’s possible that the strength of this relationship is mainly due to the preferences of the user group employed by this study; i.e., that highly technical young adults have a general preference for websites or systems that provide pragmatic or practical value. But it is also possible that the usefulness of a website is the single most important criteria for determining its overall goodness. This explanation supports the position of Greenberg and Buxton (2008), who argued that usefulness is a much more critical design goal than usability. Echoing Thomas Landauer in *The Trouble with Computers*, they argued that “most computer systems are problematic because...they do too little that is useful” (p. 116). In other words, usefulness—not usability—is the driving force behind product adoption.
However, it should also be noted that the results of the quantitative analysis also suggested that usefulness alone was not sufficient to consider a website good. On the contrary, usability, aesthetics and enjoyment all had significant relationships to goodness; the major difference is that usefulness had a consistent relationship across all systems and contexts while the other variables had different relationships depending on the system and the context. This result suggests that although users’ overall evaluation of a system is heavily influenced by usefulness, it is also multi-faceted and context-dependent. This result is consistent with the holistic notion of UX that includes all aspects of interacting with a system (Hassenzahl, 2003) and supports the view that UX should consider users’ entire interactive experience by considering both instrumental and non-instrumental goals (Bargas-Avila & Hornbæk, 2011). However, it differs from a widely held belief that UX is centered on the “non-utilitarian” aspects of interactions (Law, et al., 2009). The results of this study actually suggest the opposite: that the pragmatic aspects of usefulness provide the foundation of UX while hedonic and non-instrumental aspects (e.g., aesthetics and enjoyment) play important, but complementary, roles. Therefore, UX evaluators should broaden their perspective and probe for issues of usefulness—in terms of both pragmatic and hedonic attributes—in addition to issues related to usability and affect or emotion.

5.5 RQ4: User Definitions of Usefulness

**RQ4: What criteria do users identify when defining or describing a useful system?**

The fourth research question addressed the concept of usefulness from the users’ perspective. After completing all three scenarios and rating all three websites for usefulness, usability, aesthetics, enjoyment, and goodness, participants were asked to
respond verbally to the following open-ended question: “What makes a website useful to you? What criteria do you look for?” Therefore, RQ4 was answered by examining the results of the qualitative analysis.

As discussed previously, four themes emerged from the qualitative analysis to represent the major dimensions of usefulness from the users’ perspective. The first theme was that a useful website should be appropriate for the context in which it is used in that it helps users achieve their goal(s), provides relevant functionality, and offers relevant and accurate information. The second theme was that a useful website should be simple and easy to use. Specifically, participants stated that it should be straightforward and easy to navigate, allow for quick and efficient access to the desired information, organize content logically, and have a clean and simple interface. The third theme was that a useful website should be visually attractive with appropriate color schemes and a wellcrafted and professional-looking design. The fourth and final theme was that a useful website should provide pleasurable interactions in terms of providing familiar and consistent experiences, presenting an engaging “feel” or “flow,” being free of pop-up ads or other irritations, and offering tools for customization.

These findings suggest that the meaning of usefulness in the emerging UX Phase should be updated to go beyond the simple task-based notion of usefulness that was prevalent in the Usability Phase. Instead, usefulness must be expanded to include the pragmatic and hedonic attributes of using interactive systems (Hassenzahl, 2003) by considering, for example, the visual attractiveness of the website and whether it provides pleasurable interactions in addition to determining whether it is appropriate for the context in which it is used and whether it is simple and easy to use. One potential
objection to the assertion that usefulness consists of both pragmatic and hedonic components is that only five participants (13.9%) actually mentioned all four themes even though each theme was discussed by nearly half of all participants (or more). This finding suggests that a universal definition of usefulness may not exist because usefulness is particular to specific users in specific situations. But, a closer examination of the data shows that this objection is not entirely accurate. Although few participants mentioned all four themes together, all 36 participants mentioned one or both of appropriateness to context or simple and easy to use while 26 participants (72.2%) mentioned one or both of visual attractiveness or pleasurable interactions. This finding implies that users seem to have a universal appreciation for the pragmatic attributes of usefulness (e.g., that a website is appropriate for the context and/or is simple and easy to use) and that a majority of users also appreciate the hedonic attributes of usefulness (e.g., that the website is visually attractive and/or provides pleasurable interactions). Therefore, the qualitative data support the conclusion from the previous section that the pragmatic (instrumental) attributes of usefulness provide the foundation of UX and that hedonic (non-instrumental) attributes play a complementary, but important, role.

5.6 Summary

This chapter discussed the findings in relation to the four research questions addressed by this research. First, it was confirmed that the usefulness of a system is shaped by the context in which it is used and users’ technical background. Second, it was determined that although usability is a major element of usefulness, they are distinct concepts and require a different set of tools to understand and evaluate. Third, it was found that aesthetics has a complicated relationship with usefulness, with users seeming
to value its contribution when describing usefulness in general but discounting it when providing concrete ratings. Fourth, it was discovered that enjoyment is not related to the pragmatic dimension of usefulness (as it was defined in this study) but that some users associate usefulness with the provision of pleasurable interactions. Fifth, users’ overall evaluation of a system (i.e., its goodness) is highly influenced by perceptions of usefulness regardless of the system or the context, suggesting that the pragmatic component of usefulness provides the foundation of UX while hedonic attributes play complementary roles. Finally, users identified four major dimensions of usefulness and although few users identified all four dimensions, there is evidence that usefulness consists of both pragmatic and hedonic components.
6. CONCLUSIONS & FUTURE WORK

This chapter concludes the dissertation. First, the previous chapters are summarized, the research questions are presented, and the major findings are reviewed. Second, the findings are further examined by considering their implications for both research and practice, specifically that evaluators should 1) probe for issues related to usefulness in addition to usability, 2) broaden the scope of evaluation to include both pragmatic and hedonic attributes, and 3) vary evaluation contexts to provide diverse experiences and gain multiple perspectives. Third, three potential limitations of the study are discussed: the assumption that complex constructs can be simplified and measured quantitatively, the constraints of studying these constructs in a controlled environment, and the treatment of context as an experimental variable. Finally, remaining unanswered questions and future research directions are described, including simplifying the experimental methods, exploring perceptions of usefulness in other contexts (including other user groups, other system types, and other tasks), investigating the effects of individual elements of context to determine which elements have the greatest influence on perceptions of usefulness (e.g., task vs. social environment), updating and refining a user-based definition of usefulness, and developing novel UX evaluation methods that capture the multi-faceted and complex nature of system usefulness.

6.1 Introduction

This dissertation described the results of a study into the nature and meaning of usefulness. The concept of usefulness has implicitly been a critical goal of system evaluation, but the meaning of usefulness has changed over time from a narrow system-based focus on reliability to a broader user-centered focus on ease of use and usability.
User experience (UX), which considers both the practical and emotional aspect of system use, is emerging as the dominant research paradigm in HCI. Thus, it is likely that the meaning of usefulness needs to be updated accordingly.

Designing useful systems has long been cited as a primary goal of employing user-centered design methods, but there is surprisingly little consensus about a precise definition of the term “usefulness” and studies of usefulness have been largely absent from HCI research. An examination of the literature revealed 28 different definitions of usefulness (or utility), from which a working definition of usefulness was proposed as “the extent to which a system’s functions allow users to complete a set of tasks and achieve specific goals in a particular context of use.” This definition implies that providing functionality is not sufficient for determining a useful system; instead, it must appropriately fit the context of use. Despite this obvious connection, few studies have examined it and there exists no coherent knowledge base surrounding this relationship. Consequently, this research sought to fill this gap and both confirm and explore how the usefulness of a system is shaped by the context in which it is used. The meaning of context has been debated within HCI and elsewhere, but this research adopted a simple framework for context as the user, the task, the tool and the environment. Furthermore, it has been suggested that perceptions of usefulness may also be influenced by the usability, aesthetics, and enjoyment of the system, but there is little empirical evidence describing these relationships. Thus, the following four research questions were addressed:

**RQ1:** How and to what extent are perceptions of usefulness shaped by the context—the user, task, and environment—in which the system is used?
**RQ2:** How and to what extent do perceptions of usability, aesthetics, and enjoyment influence perceptions of usefulness?

**RQ3:** To what extent do perceptions of usefulness affect users’ overall evaluation of a system?

**RQ4:** What criteria do users identify when defining or describing a useful system?

A repeated measures experiment was designed to investigate these questions. In the experiment, 36 participants interacted with three different information portal websites in three different contextual scenarios and were then asked to rate each system in terms of usefulness, usability, aesthetics, enjoyment, and goodness (their overall evaluation of the system). There were six major findings from this study. First, it was confirmed that the usefulness of a system is shaped by the context in which it is used. Second, it was determined that although usability is a major element of usefulness, they are distinct concepts and require a different set of tools to understand and evaluate. Third, it was found that aesthetics has a complicated relationship with usefulness, with users seeming to value its contribution when describing usefulness in general but discounting it when providing concrete ratings of usefulness. Fourth, it was discovered that enjoyment is not related to the pragmatic dimension of usefulness (as it was defined in this study) but that some users associate usefulness with the provision of pleasurable interactions. Fifth, users’ overall evaluation of a system (i.e., its goodness) is highly influenced by perceptions of usefulness regardless of the system or the context, but other system attributes are also important in certain conditions. Finally, users identified four major dimensions of usefulness: appropriateness to context, simplicity and ease of use, visual
attractiveness, and pleasurable interaction. However, few users identified all four
dimensions, suggesting that there is no universal definition of usefulness. These findings
have several implications for research and practice that will be discussed in the following
sections.

6.2 Probe for Issues Related to Usefulness

According to the quantitative results, usefulness was the only system attribute that
had a consistently positive influence on users’ overall evaluation of a system. Meanwhile,
the qualitative results showed that pragmatic aspects of usefulness—a system’s
appropriateness for the context in which it is used and/or its simplicity and ease of use—
were universally identified by all participants as criteria of a useful website. These
findings suggest that the pragmatic component of system usefulness provides the
foundation for the overall User Experience (UX), with hedonic attributes (e.g., aesthetics
and enjoyment) playing a complementary role. Consequently, researchers and evaluators
who wish to evaluate interactive systems are encouraged to probe for issues related to
usefulness in addition to issues of usability in order to gain a better understanding of how
the system is perceived and interpreted by its users. To date, the vast majority of system
evaluation efforts have focused primarily on usability. The findings of this research have
confirmed the importance of evaluating systems for ease of use, with the quantitative
analysis showing a strong relationship between usability and usefulness and the
qualitative analysis demonstrating that many users (approximately 75% of study
participants) consider simplicity and ease of use as important aspects of a useful system.
Unfortunately, a focus on usability often comes at the expense of usefulness because the
task-centered nature of usability testing ignores how well the system fits into the broader
use context (Dicks, 2002). Exploring issues related to system usefulness during usability evaluation is not an uncommon practice, but it is infrequent, informal, and without a coherent connection to evaluation goals (Nørgaard & Hornbæk, 2006). To address this challenge, evaluators are encouraged to: 1) develop a shared understanding among the design and evaluation team for what usefulness means for the system being evaluated, 2) incorporate questions about system usefulness into evaluation plans, and 3) purposefully address issues of usefulness (both formally and informally) when talking with users during user testing. When doing so, evaluators are encouraged to take a broad view of usefulness by considering how well the functions provided by the system allow its users to complete their tasks in the intended use context, which includes the background of the users and their physical and social/cultural environments.

6.3 Broaden the Scope of Evaluation to Include Hedonic Attributes of Usefulness

The second implication of these findings is that usefulness is not shaped purely by pragmatic goals, and the scope of evaluation should be broadened to account for its hedonic components. As mentioned previously, the usability evaluation methods developed in the 1980’s and 1990’s are still the most widely used evaluation methods today (Hornbæk, 2010). While these methods have withstood the test of the time, their scope is limited primarily to those issues related to learnability and ease of use. The results of this research suggest that although usability is an important factor in determining users’ perceptions of usefulness, it is not the only factor. Instead, perceptions of system usefulness are also influenced by its visual appearance and by its provision of pleasurable interaction. These factors are not universal; users seem to perceive and discuss usefulness differently depending on their background, experience, or their
specific context of use. However, they should be considered a sufficient starting point for defining evaluation goals. Based on the results of this study, any evaluation effort should address the following questions:

**Result #1:** The usefulness of a system is shaped by the context in which it is used.

a. What are the primary goals of the intended users? Does the website/system help them achieve those goals?

b. Do the functions provided by the website/system assist users in achieve their goals?

c. Does the website/system provide all of the relevant content users may need in the course of achieving their goals?

**Result #2:** Although usability is a major element of usefulness, they are distinct concepts and require a different set of tools to understand and evaluate.

a. Does the website/system provide a simple and logical method of navigation? Can users get where they need to go? If they get lost, can they get back?

b. When using the website/system, can users quickly and efficiently find what they’re looking for?

c. Is the website/system organized logically? Is the amount of clutter minimized?

d. Is the website/system simple? Are the primary functions of the website straightforward and easily accessible?
Result #3: Aesthetics has a complicated relationship with usefulness, with users seeming to value its contribution when describing usefulness in general but discounting it when providing concrete ratings of usefulness.

a. Does the website/system have an attractive color scheme, contain relevant and unobtrusive graphics, and use a legible font? Is it pleasing to look at?

b. Does the website/system comply with industry standards for visual design? Does it appear professionally designed?

c. In general, is the website/system visually appealing?

Result #4: Enjoyment is not related to the pragmatic dimension of usefulness (as it was defined in this study) but some users associate usefulness with the provision of pleasurable interactions.

a. Does the website/system provide a pleasurable interactive experience?

b. Is the website/system free or almost-free of distracting or potentially distract features?

c. If applicable, does the website provide options for personalization, customization, or tailorability?

d. Does the website/system conform to users’ expectations? Is it similar to other websites in its domain?

Overall, these questions address the fifth finding of this research—that users’ overall evaluation of a system (i.e., its goodness) is highly influenced by perceptions of usefulness—by providing a framework to address all possible aspects of system usefulness. Some of these questions may be answered using traditional evaluation methods (e.g., heuristic evaluation or cognitive walkthrough) but others may require
additional methods or tools. In general, any thorough evaluation effort should address as many of these questions as possible, which may require the use of multiple methods. In other words, evaluation should be seen as an ongoing, iterative process of data collection and analysis in which evaluators continually gather user feedback from multiple perspectives and triangulate findings to ensure validity and consistency.

6.4 Vary Contexts of Evaluation

The third implication of these findings is that evaluation contexts should be varied to reflect the dynamic and unpredictable nature of system use. The results of this study have shown the importance of context in determining users’ perceptions of a system. Results from the quantitative analysis showed that users’ perceptions of usefulness were dramatically different depending on the context; for instance, the mean usefulness rating of ipl2 was 5.11 under the Climate Change scenario but 3.61 under the First Computer scenario, and the other systems followed a similar pattern. Additionally, approximately 80% of the participants mentioned that their idea of a useful system was highly dependent on the context in which they were using it. It follows that evaluators can be more sensitive to the effects of context on user perceptions and opinions by varying the contexts in which systems are evaluated. This can be done by randomly sampling a set of tasks from a master task list so that each user has a unique experience with the system; providing users with tasks of diverse complexity and difficulty so that users experience the system under different conditions; conducting evaluation studies in multiple locations to eliminate potential setting effects; allow users to modify laboratory environments to more accurately simulate their real world environments; providing users with opportunities to multi-task during the experiment to more closely mimic real world use;
recruiting users with different levels of background knowledge and technical experience to represent the diverse group of users likely to be using a system; and, allowing users the freedom to move to a different task or system if they wish. Adding any of these components to a user testing scenario would eliminate any semblance of experimental consistency, which would jeopardize the overall validity of the experiment and decrease the generalizability of its findings. However, it is unlikely that a typical usability evaluation would achieve the standards necessary for experimental validity even if none of these measures were taken.

Usability evaluations are typically conducted as follows: put users in a lab, ask them to perform certain tasks, and record the results. Throughout the process, evaluators try to adhere to a strict laboratory-style procedure in order to presumably provide valid, reliable findings (Nørgaard & Hornbæk, 2006). Unfortunately, there is substantial evidence that standard usability tests suffer from numerous threats to experimental validity, including a failure to control for the wildcard effect, instrumentation bias, and confounding errors (Gray & Salzman, 1998). Additionally, the small sample sizes—typically between 8 and 16 users (Barkhuus & Rode, 2007; Sauro & Lewis, 2009)—are inadequate for making large-scale assumptions about the overall usability of a system, particularly because these participants are rarely (if ever) randomly sampled from the population (Dicks, 2002). Furthermore, it has been shown that the outcome of a usability evaluation depends on the characteristics of the participants, the number and type of tasks included, and the skills and background of the evaluators (Lindgaard & Chattratichart, 2007). In other words, although many evaluators attempt to follow strict laboratory-style procedures, they are nearly impossible to adhere to because of practical and logistical
limitations. Since standard usability testing procedures are unlikely to achieve experimental validity, evaluators may be better served by embracing the unstructured and informal nature of laboratory-based user testing rather than striving for the unattainable promise of generalizability and reliability.

6.5 Limitations

There are three limitations of this study that need to be acknowledged. The first limitation is that this study assumed the existence of a quantitative relationship between several concepts that are not typically quantified. Study participants were asked to complete a post-scenario survey by providing numeric ratings (1 to 7) on several items which were assumed to represent usefulness, usability, aesthetics, enjoyment, and goodness. While the post-scenario survey was shown to exhibit construct validity, this only confirmed that the sub-scales could distinguish between distinct constructs. There is no guarantee that the ratings accurately reflect how these concepts are actually perceived by users in real world settings; they are merely experimental conveniences that allow for statistical analysis. In reality, perceptions of usefulness, usability, aesthetics, enjoyment, and goodness are likely more complex than a series of numeric ratings on a survey. However, these ratings are necessary in order to generate statistically valid conclusions about the relationships between these constructs. In order to capture the nuances of users’ perceptions and investigate whether the conclusions made from this study are generalizable to real world situations, future studies will supplement quantitative data with qualitative observations and interviews “in the wild” with real users of real systems.

A second limitation of the study is that the dependent variables—usefulness, usability, aesthetics, enjoyment, and goodness—were measured after one structured
interaction with a system in a controlled laboratory setting. In all cases, participants were given a contextual scenario (one of which was “no scenario”) and given just five minutes to interact with the system before rating the system on those five dimensions. While there is evidence that users form impressions within the first few seconds of interacting with a website (Lindgaard, et al., 2006), it is also likely that their perceptions are dynamic and subject to change with extended usage (Karapanos, et al., 2009). Additionally, participants’ interactions with each website were restricted to one specific task, which means that their perceptions of each website were based on a single, standalone experience (most participants had no previous experience with any of the websites).

While this restriction was a necessary byproduct of the experimental design, it is possible that the results may have been different if users had been given multiple tasks of varying complexity and difficulty. Furthermore, users were instructed to stay on the given website until either 1) they felt their task was complete or 2) the five minute time limit was reached. In cases where participants had difficulty completing a task, several participants remarked that they would normally go to a different website or move to another task if they were not in an experimental setting. By forcing them to stay on each website until they completed the task, it is possible that participants’ perceptions were influenced—negatively or positively—simply because they were forced to spend extra time trying to complete the assigned task. In order to more accurately simulate real world experiences with interactive systems and capture a more holistic perspective on system usefulness, future studies will provide multiple tasks at differing levels of complexity, supplement quantitative findings with qualitative think aloud data, and allow more freedom to participants in choosing whether to complete a task on a given website.
A third limitation of this study regards the treatment of context. In order to simplify the experimental procedure and subsequent statistical analysis, it was assumed that the effects of context can be measured in a laboratory setting by varying the details of descriptive textual scenarios. This assumption implies 1) that people can meaningfully and accurately imagine themselves in a specific situation simply from reading a short description and 2) that doing so leads to significant differences in their actions and perceptions. The assumption that context can be simulated in a laboratory setting is a common assumption throughout HCI research; putting users in a laboratory and providing them with a descriptive scenario is a common component of user testing (c.f. Scenario-Based Design, Carroll, 2000). However, many researchers have argued that context is a complex, dynamic, and often indeterminate element of system use (e.g., Bødker, 2006), and would likely object the idea that context can be controlled in an experimental setting simply by modifying the details in a textual scenario. This issue was addressed in the current study by asking users to read the scenario aloud and briefly summarize it before interacting with the system. While this strategy did force all participants to understand each scenario before beginning the task, there is no guarantee that they considered all the details of the scenario once they started interacting with the system. But, the goal of this research was not to provide an exhaustive list of contextual factors that influence users’ perceptions of usefulness. Instead, this research only intended to confirm that the usefulness of a system is inextricably linked to the context in which it is used. Future studies will provide a more complete investigation of how different contextual factors influence perceptions of usefulness. These studies may
include controlled laboratory experiments, naturalistic field studies of real world systems, or a combination of the two.

6.6 Future Work

The results of this research lead to a number of promising research directions that would further enhance our understanding of usefulness and its relation to UX. The first possible research direction is to explore ways of streamlining and simplifying the experimental methods used in this study in order to more directly test the results obtained. One possible option would be recruit a larger number of participants and adopt a strictly between-groups experimental design in which participants interact with only one contextual scenario and only one system. To reduce the sample size needed for a 3x3 between-groups experiment, future studies could focus on only one or two websites instead of three websites, or two contextual scenarios instead of three scenarios. In a shortened experiment, participants could then be asked to spend more time on each website to ensure that their perceptions are not influenced by limited exposure to the website. A shortened experiment would also allow for a more comprehensive post-scenario survey that contains either multiple rating scales for each construct to enhance construct validity or rating scales for different constructs (e.g., engagement, pleasure, flow, etc.) to more accurately reflect the complexity of UX.

The second possible research direction is supplementing the findings from this study by performing similar studies in different contexts, including different types of users, different tasks, and different types of systems. The users recruited for the current study represented a specific population: young, highly educated, and technically savvy adults. It is reasonable to assume that different user groups, such as the elderly, non-
technical research scientists, or people from low-income households, may hold different perceptions of what usefulness means and how it relates to other system attributes. Furthermore, the tasks used in this study represented two types of tasks: an imprecise information exploration task and a specific information retrieval task in two different domains (climate science and computer history). Future studies will include a greater diversity of tasks, both in terms of type and domain, to better reflect the multi-purpose nature of modern web-based applications. Finally, the systems used in this study were information portal websites. Future studies will compare systems of various types, to investigate whether, for example, users evaluate information portal websites the same way as e-commerce websites, social networking websites, or news aggregation websites. Additionally, future studies can include systems with different interaction styles, including gesture-based mobile applications and touch-based tablet or kiosk systems. This research direction can also compare perceptions of system usefulness in different settings, such as classrooms and hospitals or banks and libraries. These studies would also build on the work of van Shaik and Ling (2009; 2011) and Hartmann et al. (2007), who have already demonstrated that users’ perceptions of websites are dependent on contextual factors but did not explicitly focus on how contextual differences influenced perceptions of usefulness.

The third possible research direction is to explore whether and how individual elements of context influence user perceptions of usefulness. The results of the current study showed that user perceptions of system usefulness are shaped by the context in which it is used. But, context was treated as a single experimental variable consisting of four main elements: the user, the task, the tool, and the environment. Therefore, it is
reasonable to consider whether any of the individual elements of context have a greater effect than others. For example, do users evaluate systems differently in educational contexts compared to work-related or personal contexts? Do users in computer labs evaluate systems the same way as when they are in their home or workplace? In addition to comparing different contextual elements, this research direction can investigate the magnitude of contextual differences and how they impact user perceptions of usefulness. For example, this research direction can examine whether users perceive systems differently when tasked with winning a $20 bet or a $2000 bet, or when they need to earn a passing grade to graduate or avoid being placed on academic probation.

The fourth possible research direction is further refining the working definition of usefulness proposed in this dissertation. This definition was created from the perspective of HCI researchers, but, as shown from the qualitative analysis, it is likely that users define usefulness differently than HCI professionals. The user-based definition of usefulness presented in this dissertation was based on participants’ responses to a single open-ended question in which users spoke in general terms about their definition of a useful website. By contrast, this research direction will seek to ground users in real world examples of systems they consider useful, provide a framework for understanding and interpreting how to define usefulness, and allow for a more in-depth discussion regarding the factors and criteria that make a system useful (or not). This research direction will employ different methods (including observations, focus groups, and interviews) to gather data from multiple user groups in a diverse array of contexts in order to refine and update our understanding of what usefulness means for users of modern interactive systems.
Finally, the fifth possible research direction will explore methods for broadening the scope of evaluation to include both the pragmatic and hedonic aspects of usefulness. Since usability evaluation is still considered the “gold standard” for system evaluation, in some cases this research direction will explore ways of enhancing traditional methods to make them more appropriate for evaluating systems in modern computing contexts. In other cases, this research direction will develop and assess new methods that can be used to either evaluate multiple dimensions of usefulness or provide a more thorough understanding of one component of usefulness. In all cases, research in this direction will take advantage of the knowledge gained from the aforementioned studies of usefulness in order to develop evaluation methods that are flexible and effective in a variety of real world settings.

6.7 Summary

This chapter reviewed the main findings of this research and discussed three implications for both research and practice. First, because this research showed the importance of usefulness in determining users’ overall evaluation of a system, evaluators are encouraged to 1) develop a shared understanding of usefulness that considers how well the system fits into its intended context of use, 2) incorporate plans to assess system usefulness when defining evaluation goals, and 3) probe for issues related to usefulness when conducting evaluations with users. Second, since usefulness was shown to be a multi-dimensional construct consisting of both pragmatic and hedonic attributes, it recommended that the scope of evaluation be broadened beyond usability and usefulness to also include issues related to visual attractiveness and the provision of pleasurable interactions. Third, as this research showed how perceptions of system usefulness are
shaped by the context in which it is used, evaluators are encouraged put aside any attempts to adhere to strict laboratory style experimental procedures and vary evaluation contexts so that data can be gathered from multiple perspectives and under varying conditions. Next, three potential limitations of this research were discussed: the assumption that complex constructs can be quantified and analyzed statistically, the constraints of studying these concepts with limited time periods and restricted behavioral guidelines, and the treatment of context as an experimental variable. Finally, to address these limitations and further enhance our knowledge of usefulness and UX, five ideas for future work were discussed: simplifying and streamlining the experimental methods, performing similar studies with different users, tasks, and system types, investigating the individual effects of context, further exploring user-based definitions of usefulness, and developing and assessing alternative evaluation methods that encapsulate these findings.

To conclude, designing useful systems has always been at the core of HCI, but the meaning of the term “usefulness” has changed dramatically over time and there is no consensus about an exact definition. Because of an initial focus on making software systems easy to learn for novice users, HCI evaluation research has focused primarily on usability. But as UX research emerges as the dominant research paradigm in HCI, it becomes clear that traditional evaluation methods are no longer suitable for capturing the multi-faceted and dynamic use contexts typical of modern computing. Therefore, the field of HCI needs to re-focus its evaluation efforts on system usefulness, but this requires a thorough understanding of what it means for a modern interactive system to be useful. This dissertation research provided a first step in this direction by showing that perceptions of system usefulness are shaped by the context in which it is used, that
usability is an important factor (but not the only factor) that determines system usefulness, and that usefulness consistently has a major influence on users’ overall evaluation of a system. This research has also shown that although there is no common definition of usefulness, it is a multi-dimensional construct consisting of both practical and affective attributes which require multiple tools and methods to capture and assess effectively. The knowledge gained about usefulness can be used to inform the design and development of more robust and flexible evaluation methods that are applicable in real world settings. When armed with these powerful evaluation methods, HCI researchers and practitioners can be more confident that the systems they design will provide a high-quality user experience and positively impact people’s lives.
LIST OF REFERENCES


APPENDIX A. POST-SCENARIO SURVEY

Please indicate your level of agreement with each of the following questions.

1) Visually, everything goes together on this website.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Strongly Disagree Neutral Strongly Agree

2) The color composition is attractive.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Strongly Disagree Neutral Strongly Agree

3) The layout appears professionally designed.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Strongly Disagree Neutral Strongly Agree

4) The layout is pleasantly varied.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Strongly Disagree Neutral Strongly Agree

5) I find using the website to be enjoyable.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Strongly Disagree Neutral Strongly Agree

6) The actual process of using the website is pleasant.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Strongly Disagree Neutral Strongly Agree
7) I have fun using the website.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree | Neutral | Strongly Agree

8) The website is easy to use.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree | Neutral | Strongly Agree

9) I feel in control when I am using this website.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree | Neutral | Strongly Agree

10) The website requires little effort to use.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree | Neutral | Strongly Agree

11) Using the website is effective.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree | Neutral | Strongly Agree

12) The website provides the right functions.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree | Neutral | Strongly Agree

13) I am able to use the website to complete my task(s).

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree | Neutral | Strongly Agree

14) I am able to use the website to fulfill my goal(s).

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree | Neutral | Strongly Agree
15) The website fits my current situation.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree  Neutral  Strongly Agree

16) The website is useful.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Strongly Disagree  Neutral  Strongly Agree

17) I judge the website to be:

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Bad  Good
APPENDIX B. POST-STUDY QUESTIONNAIRE

1) Have you ever used the ipl2 website before?
   ( ) Yes
   ( ) No

2) Have you ever used the RefSeek website before?
   ( ) Yes
   ( ) No

3) Have you ever used the Awesome Library website before?
   ( ) Yes
   ( ) No

4) Please respond to each of the following statements by selecting the option that most closely reflects your own knowledge and experience regarding the Internet. There are no right or wrong answers, so please answer honestly.

<table>
<thead>
<tr>
<th>Statement</th>
<th>I don’t understand this statement and cannot respond (0)</th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither agree or disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>If a computer problem occurs while I am using the Internet, I usually know how to fix the problem.</td>
<td>( )</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>I know how to create a website.</td>
<td>( )</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>I know some good ways to avoid computer viruses.</td>
<td>( )</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>I am familiar with HTML.</td>
<td>( )</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>I know how to enable and disable cookies on my computer.</td>
<td>( )</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>I am able to download a “plug-in” when one is recommended in order to view or access something on the Internet.</td>
<td>( )</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>I understand most computer terms that have to do with the Internet.</td>
<td>( )</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
</tbody>
</table>
I can usually fix any problems I encounter when using the Internet. ( ) ( ) ( ) ( ) ( ) ( )

I help others who are learning to use the Internet. ( ) ( ) ( ) ( ) ( ) ( )

I download and install software updates from the Internet when necessary. ( ) ( ) ( ) ( ) ( ) ( )

I regularly update my virus protection software. ( ) ( ) ( ) ( ) ( ) ( )

I can design a nice background and/or signature for the e-mail messages I send. ( ) ( ) ( ) ( ) ( ) ( )

I know what a browser is. ( ) ( ) ( ) ( ) ( ) ( )

I have changed the settings or preferences on my computer that pertain to my Internet access. ( ) ( ) ( ) ( ) ( ) ( )

5) Have you taken (or are currently taking) any of the following courses? (check all that apply)

[ ] INFO 110 - Human-Computer Interaction I
[ ] INFO 215 - Social Aspects of Information Systems
[ ] INFO 310 - Human-Computer Interaction II

6) What is your major?

( ) Information Systems (BSIS)
( ) Information Technology (BSIT)
( ) Software Engineering (BSSE)
( ) BS/MS Accelerated Degree Program
( ) BS/JD Accelerated Degree Program
( ) None of the above

7) What is your class standing?

( ) Freshman
( ) Sophomore
( ) Pre-Junior
( ) Junior
( ) Senior

8) How old are you? [Dropdown: 18-65]

9) Are you:

( ) Male
( ) Female
VITA

Craig Matthew MacDonald

Education

Ph.D., Information Studies 2012
*College of Information Science & Technology, Drexel University*

M.S., Applied & Mathematical Statistics 2007
*Department of Statistics, Rutgers-The State University of New Jersey*

B.A., Mathematics-Statistics 2004
*Department of Mathematics & Statistics, The College of New Jersey*

Selected Publications


