What Does it Mean for a System to be Useful? An Exploratory Study of Usefulness

Craig M. MacDonald
Pratt Institute
144 West 14th Street, New York, NY 10011
cmacdona@pratt.edu

Michael E. Atwood
Drexel University
3141 Chestnut Street, Philadelphia, PA 19104
atwood@drexel.edu

ABSTRACT
HCI has always focused on designing useful and usable interactive systems, but usability has dominated the field while research on usefulness has been largely absent. With user experience (UX) emerging as a dominant paradigm, it is necessary to consider the meaning of usefulness for modern computing contexts. This paper describes the results of an exploratory study of usefulness and its relation to contextual and experiential factors. The results show that a system's usefulness is shaped by the context in which it is used, usability is closely linked to usefulness, usefulness may have both pragmatic and hedonic attributes, and usefulness is critical in defining users' overall evaluation of a system (i.e., its goodness). We conclude by discussing the implications of this research and describing plans for extending our understanding of usefulness in other settings.

Author Keywords
Usefulness; usability; UX; evaluation

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
The pervasiveness of the web and the blurring line between work and personal technology use has resulted in drastic changes to how, where, and why people use technology [5]. User experience (UX) is emerging as a new paradigm for HCI [2] but usability remains the "gold standard" for evaluation [23], even though usability is not the only factor that determines success; as Douglas Engelbart once reportedly remarked, "If ease of use was the only valid criterion, people would stick to tricycles and never try bicycles" [3]. Nevertheless, usability research has dominated HCI over the past 40 years, yielding a great deal of insight about how to design systems that are easy to use and easy to learn. By comparison, little effort has been spent on determining whether or not systems are *useful* [32].

DIS '14, June 21 - 25 2014, Vancouver, BC, Canada

http://dx.doi.org/10.1145/2598510.2598600

The concept of usefulness is not new to HCI, as designing useful systems has long been cited as one of the primary goals of user-centered design [15]. However, studies of usefulness are largely absent from HCI research and there is no widely accepted definition of the term. It seems reasonable to assume that the usefulness of a system depends on the context in which it is used and it is also likely that usefulness is closely related to usability and perhaps other experiential factors, but there is little empirical evidence describing these relationships.

As one of the earliest and most ardent proponents of a user-oriented perspective to computing, 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 to "actually" useful. As a first step in this direction, we will present the results of an exploratory study aimed at more closely examining the construct of usefulness by considering how it is influenced by context and how it relates to other UX attributes.

WHAT IS USEFULNESS?
The first step in understanding usefulness is developing a more precise definition of the term. While some researchers make a distinction between "utility" and "usefulness" (e.g. [36]) there is substantial overlap in how the two terms are used. Therefore, we will consider definitions of both terms to provide a more complete understanding of the concept.

Defining Usefulness
Due to our interest in understanding the concept of usefulness from an HCI perspective, we restricted our search for definitions of "usefulness" (or "utility") to only HCI journals and conference proceedings (i.e., CHI, IJHCS, Computers in Human Behavior, etc.). Our review showed that even though many authors have insisted that usefulness is an important goal and worthy of attention (e.g., [18]), there is no standard, agreed-upon definition. While our review was not exhaustive or systematic, we nevertheless identified 18 distinct definitions or uses of the terms "usefulness" or "utility" that addressed one or more of the following concepts (with sample references):

- the *functions* of the system (e.g., [15]),
- the *tasks* users are trying to complete (e.g., [33]),
- the *goals* users are trying to achieve (e.g., [36]), and
- the *context* in which the system is being used (e.g., [35]).

Accordingly, we propose a working definition of usefulness as *the extent to which a system's functions allow users to complete a set of tasks and fulfill specific goals in a particular context of use.*

Usefulness and Usability

Our working definition of usefulness closely mirrors 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" [26]. This similarity is not surprising, as many have noted the close relationship between these two concepts [18]. The exact nature of this relationship, however, is difficult to explore without a clearer understanding of what "usability" means. There is little consensus about a precise definition of usability [46], but there is broad agreement that usability refers to the effectiveness, efficiency and satisfaction with which users can complete tasks on an interface. These three criteria are included in many popular definitions of usability [26, 36] and are among the most commonly cited aspects of usability [6]. The most widely used usability metrics (task completion time, number of errors, accuracy, task completion rate, and satisfaction) also map directly onto these attributes [24]. Given the diversity of perspectives and purposes of usability research, use of these measures is surprisingly consistent [22]. Thus, this research will consider "usefulness" as a system's appropriateness for a specific context and "usability" as its effectiveness, efficiency and satisfaction within that context[1].

With the terms adequately defined, the relationship between usefulness and usability becomes more apparent. Broadly, useful systems are adopted and used by people, often regardless of usability problems, with usability improvements added later. Consider early VCRs or the Internet; although they suffered from numerous usability problems at the start, they were still highly used and became more usable over time [17]. An exception to this rule is when usability problems are so severe that users are unable to use the system [14]; in these cases, poor usability actually prevents users from taking advantage of the system's capabilities, severely limiting its usefulness.

While this relationship makes sense at a conceptual level, few studies have actually explored the connection between usability and usefulness. In one case study of an e-mail program, it was reported that although users found the program to be highly usable, they did not take advantage of its functionality because they did not find it useful [33]. In another case study, researchers discovered that users were not using a system even though they had recently overhauled the interface to make it easier to use. After further investigation, the researchers concluded that users did not find the system useful because designers misunderstood how the intended users were going to use it [29]. More recently, [1] reported that users of a large-scale enterprise planning system defined its usefulness in terms of their coping strategies for dealing with usability problems, with some users inventing workarounds to circumvent the problems and others modifying their work practice to conform to the system. Together, these results provide valuable insight into the potential relationship between usability and usefulness, but they were all case studies of a particular system and its users. No studies have attempted to empirically determine the exact nature of the relationship between usability and usefulness.

Usefulness, Usability and Perceived Usefulness

Over the last two decades, the construct of "perceived usefulness" has been widely used to study acceptance of information systems as part of the Technology Acceptance Model (TAM) [7]. The TAM has been used to evaluate over 50 different types of systems, including e-mail, groupware, voicemail, digital libraries, and electronic health systems, and perceived usefulness was highly correlated with actual system use in nearly all cases [44]. The original TAM defined perceived usefulness as "the degree to which a person believes that using a particular system would enhance his or her job performance." The latest version (TAM3) has slightly revised the definition to "an instrumental belief...regarding the benefits of using a system" but the concept is still measured by questions about whether the system "improves performance" or "enhances [users'] effectiveness" in their job [45]. In short, the TAM defines "perceived usefulness" in terms of performance-related issues – effectiveness and efficiency – that are typically assumed to be aspects of usability (particularly in how it is defined in this research). Due to the substantial overlap between TAM's perceived usefulness and the HCI concept of usability, the TAM definition of perceived usefulness was considered inappropriate for this research[2].

[1] We recognize this distinction is imperfect – in fact, some authors describe usefulness as one *component* of usability [39] – but making *some* distinction was necessary due to our limited knowledge of usefulness. We consider these definitions as starting points for developing a deeper understanding (and more refined definitions) of both concepts and their relationship.

[2] Similar concepts found in other domains (e.g., requirements engineering, software engineering) are similarly beyond the scope of this research.

Usefulness and Other UX Attributes

Usefulness and Visual Aesthetics

Although the debate between form and function has been prevalent in industrial design for several decades, the study of aesthetics in HCI was virtually non-existent up until the late 1990s. One of the most well-known studies of aesthetics in HCI reported that not only were users able to distinguish aesthetics from other system attributes, but their perceptions of aesthetics were highly correlated with perceptions of usability and with overall satisfaction, leading to the conclusion that "what is beautiful is usable" [43]. Subsequent studies have consistently shown that the aesthetic appeal of a system has some type of mediating effect on the construct of usability [8, 9, 10, 21], suggesting that usefulness may be similarly influenced. However, no studies of aesthetics and usefulness have been reported.

Usefulness and Enjoyment

In addition to usability and aesthetics, it is possible that usefulness is influenced by the emotions users feel when using a system. As HCI embraces the concept of UX and enters its third wave [5], there is an increased need to understand users' subjective experiences with technology. While there are a variety of ways to capture users' affective states [2], some common approaches are to measure the pleasure [20] or enjoyment [40] users derive from using the system. Researchers continue to find evidence that these subjective factors have a strong influence on usability, but no studies have examined their effect on usefulness.

Usefulness and Context

Context is a prominent concept in our working definition of usefulness. Although many researchers have recognized the contextual dimension of usefulness, few studies have explicitly attempted to understand it. In those studies, usefulness has been defined in terms of "mis-fits" between the system and its users [4], as a measure of "conceptual fitness" in terms of the tradeoffs between effort and benefit [25], and as a social construct created *in situ* by users [1]. While these definitions are instructive in making a clear connection between usefulness and context, they have few overlapping features. Furthermore, each of these studies only speculated about the relationship between context and usefulness; none of them sought to provide empirical data to describe how the concepts were related.

METHOD

We designed and conducted an exploratory experiment to examine the effects of context on usefulness and explore the relationships between usefulness and three other UX attributes: usability, aesthetics, and enjoyment.

Context

Although the nature and meaning of context has long been debated [11], several theories germane to the study of HCI–activity theory, situated action, distributed cognition, 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: a system can only be considered useful if it fits into the context of use. As described in the previous section, previous studies of usefulness and context were field studies of particular systems in specific settings and thus did not further our understanding of this relationship in any meaningful way. By contrast, our primary goal is to collect empirical data about usefulness and context (which no previous studies have attempted to do), which requires operationalizing the concept for an experimental setting. We recognize that a laboratory study appears to be an odd fit for studying context, but we believe such approach is an appropriate starting point for research on usefulness.

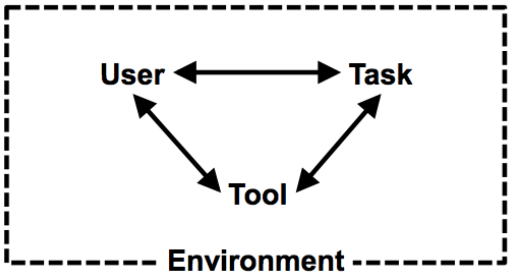

Figure 1. Four contextual factors: user, task, tool, and environment. Adapted from [41].

To begin, we examined a number of contextual frameworks and decided, for simplicity, to consider context as a combination of four factors (Figure 1): the user, the task, the tool, and the environment [41] (with the latter referring to both the physical environment and the social and cultural environment in which the interaction takes place [35]). To control for the "user" factor, we limited the study to a group of individuals (in this case, tech-savvy undergraduate students) who were likely to share common interests, knowledge, and interaction styles and/or preferences. The other three contextual factors – task, tool, and environment – were operationalized as independent variables.

Independent Variable: Task-Environment

To represent the "task" and "environment" contextual factors, we created three written scenarios: one scenario was "no scenario" and used as a control and two scenarios were similar in length and complexity but differed along the task and environment dimensions [41]. Thus, in addition to describing different physical locations (the library vs. a university dormitory room) and social situations (school research vs. a friendly wager), the scenarios also described different tasks: scenario one described an imprecise information exploration task and scenario two described a specific information retrieval task, representing "action" mode and "goal" mode, respectively [20, 40] (see Table 1).

Task-Environment 1: information exploration (action mode)	Task-Environment 2: information retrieval (goal mode)
It's 10AM on a Thursday morning. You're sitting in the 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.	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.

Table 1. The two scenarios used for this study; a third scenario was "no scenario" and used as a control.

Independent Variable: Tool

To represent the "tool" contextual factor, three educational information portals were selected: ipl2 (http://www.ipl.org), RefSeek (http://www.refseek.com), and Awesome Library (http://www.awesomelibrary.org). Information portals were chosen because they offer a range of potential use cases, are relatively easy to learn, and are the type of website many undergraduates are familiar with. The three websites selected for the study offered sufficient domain overlap but differed in other aspects, including appearance, browsing structure, and search capability. The experiment was conducted on live versions of all three websites.

Usefulness and Other UX Attributes

Since this was a laboratory study, the optimal solution was to use multi-item rating scales to measure the dependent variables. Care was taken to select rating scales that were previously validated, contained 5 items or fewer (to ease the cognitive load for participants), were distinctly related to the concepts of interest, and had minimal overlap with the other variables.

Usefulness

We were unable to locate any usefulness scales that satisfied all the above criteria. As noted previously, TAM's "perceived usefulness" was deemed inappropriate due to potential overlap with items representing usability. Instead, we developed a custom rating scale based on our working definition of usefulness as a combination of the system's functions, the users' task(s), the users' goal(s), and the use context. We did not conduct any validation studies for this scale due to the exploratory nature of the study, but we did include a fifth item – stating simply "The website is useful" – in case the four items based off the working definition did not achieve construct validity.

Usability, Aesthetics, and Enjoyment

Usability was measured with four items from the Interface Quality Scale (IQS) based on the ISO definition of usability as a measure of effectiveness, efficiency, and satisfaction [8]. Aesthetics was represented by a four-item version of the Visual Aesthetics of Websites Inventory (VisAWI), which splits aesthetics into four dimensions: simplicity, diversity, colorfulness, and craftsmanship [34; Moshagen, personal communication]. Enjoyment was measured by three items asking whether using the website is enjoyable, pleasant, and fun [40]. Finally, we included a single item for "goodness" representing "the overall evaluation (the value) of a product in a given context" [21].

Study Design

There are nine possible combinations of the *tool* and *task-environment* independent variables, but we felt it would be excessive to administer all nine treatments to each participant. Since we also wanted to avoid practice effects (i.e., using the same tool multiple times or completing the same task multiple times), we used an incomplete repeated measures design in which participants were exposed to all three levels of the *tool* variable exactly once and all three levels of the *task-environment* variable exactly once. This was achieved by creating six sets of three treatments in which each level of the two variables appeared only once (e.g., each participant would interact with each tool and complete each task a single time). Treatments were counterbalanced within each set [42], resulting in 36 possible combinations (six sets and six sequences per set). This allowed us to collect 12 observations for each of the nine *task-environment* x *tool* pairs with only 36 participants. A sample treatment set and its associated sequences are shown in Tables 2 and 3.

Procedure

The experiment took place in a usability lab. At the beginning of the experiment, participants were randomly assigned to one of the six treatment sets and presented with each treatment, one at a time, where treatments consisted of

	tool 1	tool 2	tool 3
task-environment 1	***1**	2	3
task-environment 2	4	***5**	6
task-environment 3	7	8	***9**

Table 2. Treatment set containing treatments 1, 5 and 9.

Sequence	First	Second	Third
1	1	5	9
2	1	9	5
3	5	1	9
4	5	9	1
5	9	1	5
6	9	5	1

Table 3. All six possible sequences of treatments 1, 5 and 9.

one scenario (*task-environment*) and one website (*tool*). Participants were given five minutes to think aloud while attempting to complete the scenario on the website. After each scenario, they completed a post-scenario survey consisting of 17 items (in random order) rating the dependent variables. At the conclusion of the study, they completed a post-study questionnaire.

Post-Scenario Survey

The post-scenario survey consisted of 17 items: 16 items – representing usefulness, usability, aesthetics, and enjoyment – were measured on a seven-point Likert scale (where 1 was "strongly disagree" and 7 was "strongly agree") and one item – representing goodness – was measured on a seven-point semantic differential scale with "bad" and "good" as anchors. A seven-point scale was chosen because they tend to outperform five-point scales in terms of reliability, accuracy and ease of use [12].

A confirmatory factor analysis was run in SPSS Amos, a structural equation modeling add-on to SPSS, to determine whether the scales included on the post-scenario survey actually measured the intended constructs. The data consisted of 108 observations (three observations for each of the 36 participants), which is slightly low but acceptable for a confirmatory factor analysis [19]. The initial model included 16 items from the post-scenario survey representing usefulness, usability, aesthetics, and enjoyment (goodness was not included because it only contained one item). However, the model did not achieve goodness of fit because three items were highly correlated with multiple constructs. After removing those items, the model achieved goodness of fit and exhibited construct validity[3] [19]. The 13 items from the revised model plus the single item representing goodness are shown in Table 4.

Post-Study Questionnaire

At the end of the experiment, participants completed a post-study questionnaire. First, participants were asked to respond verbally to the following question: *What makes a website useful to you? What criteria do you look for?* Next, they completed a questionnaire regarding their previous experience with the websites, their knowledge of the Internet, and their demographic information and educational background. Internet knowledge was assessed using the Internet Knowledge Measure (iKnow), a score calculated by summing participants' answers to a 14-item questionnaire regarding common knowledge and behaviors associated with Internet use [38].

Participants

Thirty-six participants were recruited for this study. As planned, our study population shared a number of common characteristics: participants' median age was 21 and all were undergraduate students (sophomore or above) enrolled at Drexel University with majors in Information Technology, Information Systems, or Software Engineering. Most were male (31; 86.1%) and reported that they had not used any of the three websites prior to completing the experiment (29; 80.6%). Nearly all participants (35; 97.2%) had taken at least one course in HCI and 24 (66.7%) had taken at least two HCI courses. Their iKnow scores ranged from 52 to 70 (the maximum possible score) with an average of 62.81, indicating average to above average Internet knowledge.

Construct / Item	Loading
Usefulness (α = .936)	
1. I am able to use the website to complete my task(s).	0.866
2. I am able to use the website to fulfill my goal(s).	0.851
3. The website fits my current situation.	0.836
4. The website is useful.	0.929
Usability (α = .879)	
5. The website is easy to use.	0.865
6. I feel in control when I am using this website.	0.839
7. The website requires little effort to use.	0.823
Aesthetics (α = .890)	
8. Visually, everything goes together on this website.	0.826
9. The color composition is attractive.	0.828
10. The layout appears professionally designed.	0.816
11. The layout is pleasantly varied.	0.809
Enjoyment (α = .915)	
12. I find using the website to be enjoyable.	0.995
13. I have fun using the website.	0.849
Goodness	
14. I judge the website to be: bad–good.	-

Table 4. The final 14 items from the post-scenario survey, their factor loadings, and the Cronbach's alpha reliability coefficient for each variable.

RESULTS

The results are presented in two parts, starting with a quantitative analysis of participants' numerical ratings followed by a qualitative analysis of participants' definitions of a useful website.

Part I: Quantitative Analysis

Because the experiment was a repeated measures experiment, the ratings provided by each participant could not be considered statistically independent observations. Therefore, the data were analyzed using a linear mixed model, a regression-based approach used to analyze repeated-measures data. This method allows for violations of the independence assumption and does not require all participants to be exposed to all possible treatments, making it ideal for incomplete repeated measures designs

[3] Detailed analyses of the model's convergent, discriminant, nomological, and face validity can be found in [31].

[47]. This study utilized a special case of mixed models called *marginal models* (also known as population-averaged models), which only model the fixed effects of the main factors on the population as a whole (i.e., they contain no random effects). Since marginal and mixed models are regression-based, the results include both overall fixed effects, which can be interpreted similarly to traditional ANOVA, and estimates of regression coefficients, which can be interpreted similarly to traditional regression analysis. Since our main interest is in understanding usefulness, we performed two separate marginal model analyses: one with usefulness as the dependent variable and one with goodness as the dependent variable (and usefulness as one covariate).

Usefulness

Using SPSS, data from the post-scenario survey were put into a marginal model with usefulness as the dependent variable. The initial model included an intercept term and fixed effects for the two main contextual factors, *task-environment* and *tool*, and four covariates: usability, aesthetics, enjoyment, and Internet knowledge (i.e., iKnow score). Goodness was not included as a covariate because it was considered an overall evaluation. The model also included interaction effects between the four covariates and the two main factors. Following a three-step top-down model-building approach [47], non-significant fixed effects were removed from the model, one at a time, until only four variables remained: usability and aesthetics (as main effects) and task-environment*aesthetics and task-environment*iKnow (as interaction effects).

To assess exactly how each fixed effect influenced usefulness, the estimated regression coefficients were examined (see Table 5). According to these estimates, higher ratings of usability were associated with higher ratings of usefulness regardless of the context, indicating a strong, positive relationship between the two variables. Interestingly, higher ratings of aesthetics were associated with *lower* ratings of usefulness, with a slightly higher (but still negative) effect under the information exploration scenario (task-environment 1). Neither task-environment nor tool had a significant fixed effect on usefulness, but task-environment had an interaction effect with users' Internet knowledge. Specifically, higher iKnow scores were associated with *lower* ratings of usefulness under the information exploration scenario (task-environment 1) but *higher* ratings of usefulness under the control scenario. Participants' iKnow score had no effect on ratings of usefulness under the information retrieval scenario (task-environment 2), likely because half of the participants provided lower usefulness ratings in this scenario.

Variable	Est. (b)	Std. Error	Sig.
usability	1.386	0.082	< 0.001
aesthetics	-0.633	0.113	< 0.001
aesthetics * task-environment 1	0.472	0.131	0.001
iKnow * task-environment 1	-0.025	0.008	0.004
iKnow * control (task-env.)	0.015	0.006	0.021

Table 5. Significant regression coefficients for the marginal model with usefulness as the dependent variable. All other coefficients can be assumed to equal zero.

Goodness

Next, the data were put into a marginal model with goodness as the dependent variable and usefulness added as a covariate. The same model-building approach was used, yielding a final model with nine variables, including three main effects (tool, usefulness, and usability) and six interaction effects (task-environment*usefulness, task-environment*aesthetics, task-environment*enjoyment, tool*iKnow, tool*usability, and tool*aesthetics).

According the estimated regression coefficients (see Table 6), usefulness was the only variable to have a consistent, positive effect on goodness, with higher ratings of usefulness associated with higher ratings of goodness regardless of the contextual factors. The effect was somewhat weaker under the information retrieval scenario (task-environment 2), but higher perceptions of usefulness were consistently related to higher perceptions of goodness.

By contrast, the effects of usability were entirely mediated by the tool being used: higher ratings of usability were associated with higher ratings of goodness for ipl2 (tool 1) and RefSeek (tool 2) but appeared to have no effect on perceptions of Awesome Library (tool 3).

Participants' Internet knowledge had an interesting effect on their ratings of goodness, as those with higher iKnow

Variable	Est. (b)	Std. Error	Sig
tool 2 (RefSeek	3.590	1.403	0.013
tool 3 (Awesome Library)	-3.233	1.270	0.013
usefulness	0.574	0.105	< 0.001
usefulness * task-environment 2	-0.408	0.111	0.001
aesthetics * task-environment 1	0.630	0.137	< 0.001
aesthetics * task-environment 2	0.855	0.150	< 0.001
enjoyment * task-environment 1	-0.379	0.123	0.003
enjoyment * control (task-env.)	0.384	0.135	0.006
iKnow * tool 2	-0.043	0.019	0.027
iKnow * tool 3	0.055	0.021	0.012
usability * tool 1 (ipl2)	0.803	0.168	< 0.001
usability * tool 2	0.551	0.196	0.007
aesthetics * tool 1	-0.736	0.180	< 0.001
aesthetics * tool 2	-0.750	0.227	0.001

Table 6. Significant estimated regression coefficients for the marginal model with goodness as the dependent variable. All other coefficients can be assumed to equal zero.

scores provided *lower* goodness ratings to RefSeek but *higher* goodness ratings to Awesome Library (iKnow had no impact on goodness ratings of ipl2). Additionally, the effects of aesthetics and enjoyment on goodness varied depending on the tool and/or the task-environment.

Part II: Qualitative Analysis

To supplement the quantitative results, data from the open-ended question, "What makes a website useful to you? What criteria do you look for?" were split into 159 data points (each representing a distinct idea). The constant comparative method [13] was used to generate a "bottom-up" definition of usefulness from the users' perspective. The coding scheme consisted of 14 codes which were then described and grouped into four themes (see Table 7): *appropriateness for context*, *simplicity and ease of use*, *visual attractiveness*, and *pleasurable interaction*s.

The most common theme (77.8%) was that a useful website is *appropriate for the context* in which it is used. Participants noted that a website should fit their goal (e.g., "if it satisfies [my] purpose"), offer relevant features or functions (e.g., filtering results, advanced search), and provide the right content (e.g., "[it] needs to be credible").

The second most common theme (75.0%) was that a useful website is *simple and easy to use*. For the participants, this meant that a website is easy to navigate (e.g., "I don't have to do a lot of work"), allows for quick access to information (e.g., "I want what I want and then I want to move on"), be logically organized and without clutter (e.g., "[I] can follow it through and not be confused"), and offer a clean, simple interface (e.g., "laid out in a way that people can see it").

The third theme (47.2%) was that a useful website has a *visually attractive* user interface (e.g., an eye-catching color scheme, appropriate graphics, readable and attractive fonts) that appears well crafted and professionally designed (e.g., "it looks well put together").

The fourth and final theme (47.2%) was that a useful website provides *pleasurable interactions*. There are a number of ways this can be done, from offering a sense of familiarity (e.g., "it looks like Google"), having a natural flow (e.g., "what kind of feel [it] gives you"), eliminating intrusive ads, or providing tools for customization.

Theme / Codes	Participants
Appropriateness for Context (n = 28; 77.8%)	
Suitable for Purpose/Goal	17 (47.2%)
Right Functionality	13 (36.1%)
Appropriate Content	10 (27.8%)
Simplicity and Ease of Use (n = 27; 75.0%)	
Easy to Use/Navigate	16 (44.4%)
Speed/Efficiency in Use	11 (30.6%)
Organized/Uncluttered	10 (27.8%)
Streamlined/Simple Design	8 (22.2%)
Visual Attractiveness (n = 17; 47.2%)	
Pleasing to the Eye	10 (27.8%)
Craftsmanship	6 (16.7%)
General Attractiveness	5 (13.9%)
Pleasurable Interaction (n = 17; 47.2%)	
Familiarity	6 (16.7%)
"It" Factor	6 (16.7%)
Irritation-Free	5 (13.9%)
Customizability	2 (5.6%)

Table 7. The four themes identified by participants. (Note: Percentages are out of 36; the percentages do not add up to 100% because participants could address multiple themes.)

DISCUSSION

Four main conclusions can be drawn from the quantitative and qualitative results of this exploratory study: a system's usefulness is shaped by the context in which it is used, usability is closely linked to usefulness, usefulness may have both pragmatic and hedonic attributes, and usefulness is critical in defining users' overall evaluation of a system.

Usefulness is Shaped by Context

It is not a surprising finding that some systems are more appropriate for certain contexts than others; for instance, the website of a sporting goods store would obviously not be as useful as a travel website if you need to make flight and hotel reservations. But in this experiment, the two scenarios (excluding the control) presented two types of information tasks and all three websites provided access to similar types of information resources. Yet still, the quantitative analysis showed a significant effect of contextual factors on ratings of usefulness and many participants defined a useful website in terms of whether it provides the "right" functions and provides access to the "right" information, where "right" is determined entirely by the user's background, purpose and goal in visiting the website. Together, these findings confirm our earlier assumption that the usefulness of a system cannot be separated from the context in which it is used.

Usability is Closely Linked to Usefulness

The quantitative analysis showed that participants who perceived a website as highly usable also perceived it as highly useful, and the major components of usability–efficiency (speed), effectiveness (goal/purpose), and satisfaction (irritation-free/"it" factor)–all emerged from the qualitative analysis as commonly cited criteria defining a useful website. While this result was not particularly surprising due to how the two terms were defined in this study, the strength of the qualitative results (75% of participants mentioned issues related to usability) suggests that this is not merely a byproduct of using these definitions but rather a confirmation that usability is a critical factor in defining a useful website. Since the two constructs are so closely linked, it is reasonable to consider whether the construct of usefulness is worth considering at all; after all, if users perceived a usable website as a useful website, why should we bother with usefulness?

The answer is simple: because usability is important but it isn't enough [27]. We've shown that usability is a critical factor and we strongly believe it should always be addressed, but usability alone is not a sufficient evaluation goal because today's highly sophisticated users have expectations that go beyond ease of use and learnability. In addition, usability evaluation methods are poorly suited for capturing issues related to usefulness [17] because they focus on task-centered user performance rather than a system's fit to context and because usefulness is typically determined by *more* than just its usability. Still, due to the critical importance of these factors to HCI, the exact nature of the relationship between usefulness and usability warrants further study.

Usefulness May Have Pragmatic and Hedonic Attributes

The quantitative analysis showed that users may actually perceive a website as less useful if they also perceive it as attractive. By contrast, the qualitative analysis showed that visual attractiveness was cited by almost 50% of participants in their definition of a useful website. These conflicting results provide further evidence of the "beauty dilemma" in which users rarely cite beauty when justifying product decisions even though it is a valued aspect of design [10]. In this study, participants were asked to make a quantified aesthetic judgment as a series of numeric ratings; since beauty is rarely experienced in this manner, it's possible that participants de-valued aesthetics out of a desire to appear rational. However, when asked to describe a useful website, they seemed more comfortable talking about aesthetics in vague or ambiguous terms (e.g., "you don't want it to be hard on the eyes"). It's also notable that many participants justified their opinion by explaining that aesthetics made the website appear easier to use or lent the website a sense of credibility or professionalism.

The quantitative analysis also showed that users who found systems useful did not also find them enjoyable, but other experiential factors were mentioned by almost half of participants as an important aspect of a useful website. There are two possible explanations for these results. First, the questions measuring enjoyment on the post-scenario survey asked whether using the website was "fun" and "enjoyable" but when answering these questions many participants explained that they associated those terms with leisurely activities like playing video games or spending time with friends. Not surprisingly, no participants used "fun" or "enjoy" when describing a useful website, instead referring to features they liked or disliked. Second, the non-significant quantitative result could be a byproduct of the exploratory experimental design, as the educational nature of the websites, the relatively mundane tasks provided in the scenarios, or the limited time frame of the experiment could have limited participants' ability to have a pleasurable experience.

In considering these results, it is possible that there is hedonic component of usefulness but it may be more nuanced than simply whether they find it attractive or enjoyable. Future studies with more interactive websites, more engaging scenarios, and more prolonged exposure to the websites may provide additional insight into whether and how perceptions of usefulness are influenced by hedonic factors.

Usefulness is a Critical Factor of Goodness

The quantitative analysis showed that usefulness was the only variable with a significant effect on ratings of goodness regardless of the contextual factors. Essentially, websites that are perceived as useful are also perceived as good (and websites perceived as less useful are perceived as less good). It's possible that the strength of this relationship is mainly due to the preferences of this particular user group, i.e., highly technical undergraduates may have a preference for systems that provide pragmatic value. But it is also possible that usefulness is the primary criteria for determining a system's overall goodness. Of course, the results also suggested that usefulness alone was not sufficient to consider a system good: the quantitative analysis showed that usability, aesthetics and enjoyment all had significant relationships to goodness depending on the context. So, although users' overall evaluation is heavily influenced by usefulness, it is also multi-faceted and context-dependent, which is consistent with notion that UX is subjective and includes both instrumental and non-instrumental components [2].

CONCLUSIONS & FUTURE WORK

This paper described an exploratory laboratory study of the concept of usefulness. The results showed that 1) the usefulness of a system is shaped by the context in which it is used; 2) usefulness is closely linked to usability; 3) usefulness may contain both pragmatic and hedonic components; and 4) users' overall evaluation of a system (i.e., its goodness) is highly influenced by its usefulness, although other attributes are important in certain situations.

These findings have two major implications for research and practice. First, due to the importance of usefulness in determining perceptions of goodness, evaluators could benefit from probing for issues related to usefulness (in addition to issues of usability) to gain a better understanding of how users perceive the system [28]. Exploring issues related to usefulness during usability evaluation is not an uncommon practice, but it is infrequent, informal, and without a coherent connection to evaluation goals [37]. Thus, we encourage evaluators to incorporate questions about system usefulness into evaluation plans and purposefully address issues of usefulness (both formally and informally) during user testing.

Second, the results of this study have demonstrated the importance of context in determining users' perceptions of a system (e.g., in addition to usefulness, context also influenced usability, aesthetics, and enjoyment). This suggests that varying evaluation contexts may be valuable

in capturing the dynamic and unpredictable nature of using technology. This can be done by, for example, randomly sampling a set of tasks from a master task list, or allowing users to modify testing environments to more accurately simulate the real world. Adding any of these components would obviously jeopardize the 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 because of practical and logistical limitations [37], confounding errors and instrumentation bias [16], or due to variation in the characteristics of participants, the number and type of tasks included, and the skills and background of the evaluators [30]. Thus, evaluators may be better served embracing the unstructured and informal nature of evaluation rather than striving for validity and reliability.

Limitations and Future Work

There are several limitations of this study that we hope to address with future research. Specifically, a laboratory study with a tightly controlled group of mostly male tech-savvy undergraduate students and a highly specific type of interface is a severe limitation on the generalizability of our results. Therefore, we plan to extend this research by employing a variety of methods (including observations and interviews) to gather data from multiple user groups in diverse environments (outside of the laboratory) to help refine our working definition of usefulness, broaden our understanding of context, and deepen our knowledge of the relationship between these concepts. Concurrently, we plan to strengthen, streamline and simplify our laboratory experiments by, for example, recruiting a larger number (and different types) of participants, focusing on fewer websites, isolating specific contextual factors, and using different types of tasks, settings, and systems.

To conclude, this research was a first step in deepening our understanding of usefulness but more work remains. We hope this exploratory study inspires other researchers to continue our efforts and develop alternative approaches to studying this critical concept.

ACKNOWLEDGEMENTS

We thank Susan Weidenbeck, Michelle Rogers, Denise Agosto, and Kasper Hornbæk for their guidance in shaping and directing this research.

REFERENCES

1. Abdelnour-Nocera, J., Dunckley, L., and Sharp, H. An Approach to the Evaluation of Usefulness as a Social Construct Using Technological Frames. *Int. Journal of Human-Computer Interaction 22*, 1/2 (2007), 153-172.
2. Bargas-Avila, J. A., and Hornbæk, K. Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. In *Proc. CHI 2011*, ACM Press (2011), 2689-2698.
3. Beale, R. Slanty Design. *CACM 50*, 1 (2007), 21-24.
4. Blandford, A., Green, T.R.G., Furniss, D., and Makri, S. Evaluating system utility and conceptual fit using CASSM. *Int. Journal of Human-Computer Studies 66*, 6 (2008), 393-409.
5. Bødker, S. When Second Wave HCI meets Third Wave Challenges. In *Proc. NordiCHI 2006*, ACM Press (2006), 1-8.
6. Chen, Y.-H., Germain, C.A., and Rorissa, A. An analysis of formally published usability and web usability definitions. In *Proc. ASIS&T 2009*, ASIS&T (2009), 1-18.
7. Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly 13*, 3 (1989), 319-340.
8. De Angeli, A., Hartmann, J., and Sutcliffe, A. The Effect of Brand on the Evaluation of Websites. In *Proc. Interact 2009*, Springer (2009), 638-651.
9. De Angeli, A., Sutcliffe, A., and Hartmann, J. Interaction, Usability and Aesthetics: What Influences Users' Perceptions? In *Proc. DIS 2006*, ACM Press (2006), 271-280.
10. Deifenbach, S., and Hassenzahl, M. The "Beauty Dilemma": Beauty is Valued but Discounted in Product Choice. In *Proc. CHI 2009*, ACM Press (2009), 1419-1426.
11. Dourish, P. What We Talk About When We Talk About Context. *Personal and Ubiquitous Computing 8*, 1 (2004), 19-30.
12. Finstad, K. Response interpolation and scale sensitivity: evidence against five-point scales. *Journal of Usability Studies 5*, 3 (2010), 104-110.
13. Glaser, B.G., and Strauss, A.L. *The discovery of grounded theory: Strategies for qualitative research*. Aldine de Gruyter, Hawthorne, NY, 1967.
14. Goodwin, N.C. Functionality and Usability. *Comm. of the ACM 30*, 3 (1987), 229-233.
15. Gould, J.D., and Lewis, C. Designing for Usability: Key Principles and What Designers Think. *Comm. of the ACM 28*, 3 (1985), 300-311.
16. Gray, W. D., and Salzman, M. C. Damaged merchandise? A review of experiments that compare usability methods. *Human-Computer Interaction 13*, 3 (1998), 203-261.
17. Greenberg, S., and Buxton, B. Usability evaluation considered harmful (some of the time). In *Proc. CHI 2008*, ACM Press (2008), 111-120.
18. Grudin, J. Utility and usability: research issues and development contexts. *Interacting with Computers 4*, 2 (1992), 209-217.

19. Hair, J.F., Black, W.C., Babin, B.J., and Anderson, R.E. *Multivariate Data Analysis (7th Ed.)*. Prentice Hall, Upper Saddle River, N.J., USA, 2009.
20. Hassenzahl, M. The Thing and I: Understanding the Relationship Between User and Product. In Blythe, M.A., Monk, A.F., Overbeeke, K, Wright, P.C. (Eds.) *Funology*, Kluwer Academic Publishers (2003), 31-42.
21. Hassenzahl, M., and Monk, A. The Inference of Perceived Usability From Beauty. *Human-Computer Interaction 25*, 3 (2010), 235-260.
22. Hornbæk, K. Current practice in measuring usability: Challenges to usability studies and research. *Int. Journal of Human-Computer Studies 64*, 2 (2006) 79-102.
23. Hornbæk, K. Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology 29*, 1 (2010), 97-111.
24. Hornbæk, K., and Law, E.L. Meta-analysis of correlations among usability measures. In *Proc. CHI 2007*, ACM Press (2007), 617-626.
25. Hsi, I. Measuring the conceptual fitness of an application in a computing ecosystem. In *Proc. WISER 2004*, ACM Press (2004), 27-36.
26. ISO 9241-11. Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs)-Part 11: Guidance on Usability (1998).
27. Jónsdóttir Johannessen, G. H., and Hornbæk, K. (2014). Must evaluation methods be about usability? Devising and assessing the utility inspection method. *Behaviour & Information Technology 33*, 2 (2014), 195-206.
28. Juurmaa, K., Pitkänen, J., and Riihiaho, S. Visual walkthrough as a tool for utility assessment in a usability test. In *Proc. HCI 2013*, BCS (2013).
29. Keil, M., Beranek, P.M., and Konsynski, B.R. Usefulness and ease of use: field study evidence regarding task considerations. *Decision Support Systems 13*, 1 (1995), 75-91.
30. Lindgaard, G., and Chattratichart, J. Usability Testing: What Have We Overlooked? In *Proc. CHI 2007*, ACM Press (2007), 1415-1424.
31. MacDonald, C. M. *Understanding usefulness in human-computer interaction to enhance user experience evaluation*. PhD dissertation, Drexel University, 2012.
32. MacDonald, C. M. and Atwood, M. E. Changing perspectives on evaluation in HCI: Past, present, and future. In *Ext. Abstracts CHI 2013*, ACM Press (2013), 1969-1978.
33. Maryniak-Nelson, J.A., and Caldwell, B.S. Experience, Utility, and Situational Appropriateness: How Does Organizational Context Influence Usability of Electronic Communications Media? In *Proc. HFES 1992*, HFES (1992), 876-880.
34. Moshagen, M., and Thielsch, M.T. Facets of visual aesthetics. *Int. Journal of Human-Computer Studies 68*, 10 (2010), 689-709.
35. Nardi, B. Studying Context: A Comparison of Activity Theory, Situated Action Models, and Distributed Cognition. In Nardi, B. (ed.) *Context and Consciousness: Activity Theory and Human-Computer Interaction*, MIT Press, Cambridge (1996), 69-102.
36. Nielsen, J. *Usability Engineering*. Morgan Kaufman, San Francisco, CA, USA, 1993.
37. Nørgaard, M. and Hornbæk, K. What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing. In *Proc DIS 2006*, ACM Press (2006), 209-218.
38. Potosky, D. The Internet knowledge (iKnow) measure. *Computers in Human Behavior 23*, 6 (2007), 2760-2777.
39. Rubin, J. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley, New York, NY, USA, 1994.
40. van Schaik, P., and Ling, J. An integrated model of interaction experience for information retrieval in a Web-based encyclopedia. *Interacting with Computers 23*, 1 (2011), 18-32.
41. Shackel, B. Usability - Context, Framework, Definition, Design and Evaluation. In Shackel, B., Richardson, S.J. (eds.) *Human Factors For Informatics Usability*, Cambridge University Press (1991), 21-31.
42. Shaughnessy, J.J., Zechmeister, E.B., and Zechmeister, J.S. *Research Methods in Psychology (9th Ed.)*. McGraw Hill, New York, NY, USA, 2012.
43. Tractinsky, N., Katz, A. S., and Ikar, D. What is beautiful is usable. *Interacting with Computers 13*, 2 (2000), 127-145.
44. Venkatesh, V., Davis, F. D., and Morris, M. G. Dead or Alive? The Development, Trajectory and Future of Technology Adoption Research. *Journal of the Assoc. for Information Systems 8*, 4 (2007), 267-286.
45. Venkatesh, V., and Bala, H. Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences 39*, 2 (2008), 273-315.
46. van Welie, M., van der Veer, G.C., and Eliëns, A. Breaking down Usability. In *Proc. Interact 1999*, IOS Press (1999), 613-620.
47. West, B.T., Welch, K.B., and Galecki, A.T. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman Hall/CRC Press, Boca Raton, FL, USA, 2007.