MIS Quarterly

RESEARCH ARTICLE

WEB AND WIRELESS SITE USABILITY: UNDERSTANDING DIFFERENCES AND MODELING USE[1]

By: **Viswanath Venkatesh**
Department of Information Systems
Sam M. Walton College of Business
University of Arkansas
Fayetteville, AR 72701
U.S.A.
vvenkatesh@vvenkatesh.us

V. Ramesh
Information Systems Department
Kelley School of Business
Indiana University
1309 E. Tenth Street
Bloomington, IN 47405
U.S.A.
venkat@indiana.edu

Abstract

Recent research has presented a conceptualization, metric, and instrument based on Microsoft Usability Guidelines (MUG; see Agarwal and Venkatesh 2002). In this paper, we use MUG to further our understanding of web and wireless site use. We conducted two empirical studies among over 1,000 participants. In study 1, conducted in both the United States and Finland, we establish the generalizability of the MUG conceptualization, metric, and associated instrument from the United States to Finland. In study 2, which involved longitudinal data collection in Finland, we delved into an examination of differences in factors important in determining web versus wireless site usability. Also, in study 2, based on a follow-up survey about site use conducted 3 months after the initial survey, we found support for a model of site use that employs the MUG categories and subcategories as predictors. The MUG-based model outperformed the widely employed technology acceptance model both in terms of richness and variance explained (about 70 percent compared to 50 percent).

Keywords: Usability guidelines, modeling, wireless

Introduction

Usability has been studied extensively in the human-computer interaction (HCI) literature (see Nielsen 2000). However, its application in the information systems literature is rather limited. The application of usability to study use can provide an important alternative perspective relative to the psychological models typically employed in IS (see Venkatesh et al. 2003). Further, usability-based models can be more directly tied to system design characteristics. Recently, a new conceptualization, metric, and measurement instrument of usability put forth by Agarwal and Venkatesh (2002) shows promise in terms of providing a holistic view of site usability. Their work employed the Microsoft Usability Guidelines (MUG; see Keeker 1997) and targeted web sites, and was tested via an application to web sites among customers and investors. We use MUG and extend prior research in three important ways: (1) examining its generalizability; (2) understanding web versus wireless site[2] usability; and (3) predicting site use.

[1]Allen Lee was the accepting senior editor for this paper. Patrick Chau was the associate editor. Moez Limayem, James Jiang, and Virpi Tuunainen served as reviewers.

[2]In this paper, the term *web site* will refer to sites accessed through traditional devices such as personal computers and laptops, and the term *wireless site* will refer to sites customized for access from wireless, handheld devices such as cell phones and PDAs.

Establishing generalizability is an important scientific goal in its own right (Lee and Baskerville 2003). Here, its importance is tied to the pursuit of the latter two objectives and the fact that we intended to collect data in support of those objectives in Finland. Hence, we had to establish the generalizability of the MUG conceptualization, metric, and instrument (objective #1) in the setting of the study (i.e., Finland) since the MUG conceptualization, metric, and instrument are new and established only in the United States (see Agarwal and Venkatesh 2002). Establishing its generalizability outside the original setting is an important step (Lee and Baskerville 2003) and allows us to ensure the existence of a common frame of reference across the original setting of the study (i.e., the United States) and the new setting (i.e., Finland).

Understanding wireless usability is important to researchers and practitioners alike. The latter part of the 1990s saw a significant shift in the way business was conducted. Business using the web became a core component of many successful organizations. A critical aspect of doing successful business on the web is the design of the web site (see Lohse and Spiller 1999) and, therefore, its usability. More recently, mobile commerce (i.e., m-commerce), which can be broadly described as applications and services accessible through a wireless, handheld device, is seen as a key technological innovation that could provide a similar business impact given the proliferation of wireless, handheld devices such as cell phones and PDAs (Osborn 2003; Sadeh 2002). Recent studies identified a key stumbling block for wireless sites as being the lack of usability (Jackson et al. 2002; Ramsay and Nielsen 2000; see also Yankee Group 2002). Understanding how wireless sites should be designed differently from web sites is still at a nascent stage. The importance of understanding the differences between web and wireless sites, our second objective, is underscored by the analogous situation in the early days of web sites, when some web sites used the implementation practice of simply porting standard paper-based marketing materials (i.e., brochureware) to their web sites and such a practice produced very poor results (Nielsen 2000; Settles 2002). Similarly, directly applying the principles of good web site design to wireless site design may not produce favorable outcomes. In sum, as the number of wireless sites and handheld devices that can access them has grown, understanding user perceptions of wireless site usability is important.

The final objective of our work is to develop and test a theoretical model of site use using the MUG conceptualization. A model based on usability would provide a rich understanding on site use that is tied to design characteristics and help provide guidance to practitioners. We develop, test, and benchmark a model of site use that employs the various categories and subcategories of MUG as predictors. Specifically, we integrate the various categories and subcategories outlined in MUG as determinants of a key system success variable (i.e., use) and incorporate web versus wireless as a moderator of key relationships in the model. In order to establish the theoretical and empirical contribution of such a model, we will benchmark our model against the widely employed technology acceptance model (TAM; Davis et al. 1989). Such a benchmarking comparison is particularly important since the dependent variable examined has been studied quite extensively in prior IS research (see Venkatesh et al. 2003). As we will argue and demonstrate, the model proposed here provides a richer and better understanding of site use. In addition to the richness of the proposed model, another strength of the current model is that while the prediction of use based on TAM and related models is robust and mature, our model represents a significant theoretical advance because of its design-centric focus (see Benbasat and Zmud 2003) and, at least in this study, its greater explanatory ability.

In sum, the current research has three major objectives:

1. Examine the generalizability of the MUG conceptualization, metric, and instrument from the United States (Agarwal and Venkatesh 2002) to a new setting (i.e., Finland).

2. Examine the differences between web and wireless sites in terms of user weights placed on various contributors to usability.

3. Develop and test a model of site use based on MUG categories and subcategories, and benchmark the proposed model against TAM.

Background: Microsoft Usability Guidelines

The International Organization for Standardization (ISO) definition of usability is "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (Karat 1997, p. 34). MUG provides a conceptualization of usability (Agarwal and Venkatesh 2002; Keeker 1997) comprising five specific categories (content, ease of use, promotion, made-for-the-medium, and emotion) that contribute to usability of a web site. Further, 4 of the 5 categories are broken down into a total of 14 subcategories. We present a brief overview of MUG categories and subcategories, key aspects of the conceptualization, metric, and instrument, and major findings.[3] Table 1 presents the list of categories and subcategories and their definitions from Agarwal and Venkatesh (2002). It should be noted that Agarwal and Venkatesh's work on MUG was focused on the study of web sites and that we aim to extend its generalizability to a new technological context (i.e., wireless sites).

The procedure employed by Agarwal and Venkatesh required users to assign weights and ratings to each of the categories and subcategories. Each user was assigned a specific role (e.g., customer) in a specific industry (e.g., airline). Employing a constant sum method, the user was asked to distribute 100 points to the various categories based on the importance of each of the categories in determining

[3]The interested reader is referred to the original article for full details.

Table 1. Microsoft Usability Guidelines: Categories, Subcategories, and Definitions

Category	Subcategory	Definition from Agarwal and Venkatesh (pp. 171-172)
Content		"…assesses the informational and transactional capabilities of a Web site"
	Relevance	"…relating to the pertinence of the content to the core audience"
	Media use	"…signifying the appropriate use of multimedia content"
	Depth and breadth	"…examining the appropriate range and detail of topics"
	Timely/current information	"…extent to which a Web site's content is current"
Ease of use		"…relates to the cognitive effort required in using a Web site"
	Structure	"…focusing on the organization of the site"
	Goals	"…relating to clear and understandable objectives"
	Feedback	"…extent to which the Web site provides information regarding progress to the user"
Promotion		"…captures the advertising of a Web site on the Internet and other media"
Made-for-the-medium		"…relates to tailoring a Web site to fit a particular user's needs"
	Personalization	"…reflecting the technology-oriented customization of the Web site"
	Community	"…capturing if the Web site provides users with an opportunity to be part of online group"
	Refinement	"…relating to the particular prominence given to current trends"
Emotion		"…taps into affective reactions invoked by a Web site"
	Challenge	"…captures the idea of difficulty"
	Plot	"…relates to how the site piques the user's interest"
	Character strength	"…relates to the credibility conveyed by the site"
	Pace	"…extent to which the site provides users an opportunity to control the flow of information"

usability of web sites in that particular industry, with a greater number of points being assigned to the more important categories. The weights assigned to each category were then distributed across the subcategories. Thus, each user assigned weights that governed the importance of the various categories and subcategories for *all sites* in that *industry* for that *role* (e.g., customer, investor) for *himself/herself* (i.e., weights are industry-specific and site-independent). An aggregation across users provided the mean and standard deviations for the weights that governed a *particular role* for *all sites* in that *particular industry*.

The user then interacted with web sites in the assigned industry and rated each web site on each subcategory using a 10-point scale. The categories were not rated since category weights were distributed among the subcategories. Each subcategory rating is site-specific and is independent of the weight assigned to the subcategory. A particular user's weighted rating for a site is then calculated by multiplying the user's weight for the subcategory by the user's rating of the particular site on that particular subcategory. The overall usability of each web site was calculated by determining the sum of the weighted ratings. Thus, each user evaluation of a site provided a holistic measure of usability for *each site* in a *particular industry* as evaluated by that *particular user* assigned a *specific role*. An aggregation of the ratings across *users* provided the mean and standard deviations for the ratings for a *particular role* for a *particular site*.

Agarwal and Venkatesh validated the metric and instrument in four industries (airline, bookstore, auto manufacturer, and car rental) with five or six web sites in each industry using two roles (i.e., customer and investor). They reported results about weights at the category level. Content was the most important category in all cases. Ease of use was also important but the pattern varied (i.e., customers found ease of use to be more important than investors). Promotion was only marginally important, and more so to investors. Made-for-the-medium and emotion were found to be interactions of product (industry) and role (customer versus investor). The importance of emotion was greater in purchases that had higher involvement. In terms of ratings, they found that the same site could receive different ratings depending on the role. For example, some sites that were market leaders in usability ratings by customers performed much poorer in the investors' eyes.

Generalizability

The first objective of this paper is to examine the generalizability of the MUG conceptualization, metric, and associated instrument from the United States (see Agarwal and Venkatesh 2002) to a new setting (i.e., Finland). Lee and Baskerville (2003) highlight the importance of establishing generalizability of a theoretical concept to different settings (see also Cook and Campbell 1979). In the context of the MUG conceptualization, metric, and instrument, Agarwal and Venkatesh established the generalizability to some extent since their study examined web sites across a range of industries using different samples of users in the United States. Since the primary focus of our theoretical extension to MUG involves data collection in Finland, we needed to establish a *common frame of reference* across settings (i.e., the United States and Finland) to ensure consistent interpretation of words, symbols, expressions, metaphors, etc. and accurate measurement (Ghorpade et al. 1999; Van de Vijver and Leung 1997). This is possible only through careful replication of an instrument or study in the new setting, eliciting feedback from participants, and examining and reconciling statistical differences via an explanation.

Given the holistic focus of the usability metric and the careful development of the associated instrument, we expect that the conceptualization and instrument would generalize from the United States to the new setting (i.e., Finland). Hence, we replicate the Agarwal and Venkatesh study in the United States by examining a previously studied industry (i.e., airline) and compare the results with those obtained by applying the MUG instrument in Finland in the same industry. To ensure that a common frame of reference and understanding of the various categories and subcategories does exist, we obtained and examined detailed justifications from the respondents for the weights and ratings assigned.

Web Sites Versus Wireless Sites

Having established the applicability of the MUG instrument in a setting outside North America, the second objective of this work is to further our understanding of wireless site usability, and the differences between web site and wireless site usability. In Agarwal and Venkatesh's (2002) study, participants accessed web sites designed primarily for access using a PC. Given that MUG provides a comprehensive and holistic conceptualization of usability using categories and subcategories that are independent of information access devices, it is reasonable to expect that the categories and subcategories in MUG can also be applied to wireless sites. The final objective of this work is to present a model of web and wireless site use based on MUG.

We extend the work of Agarwal and Venkatesh from a focus on categories only to fully consider the various subcategories that they presented. As noted earlier, although MUG has 5 categories, 4 of the 5 categories are further broken down into subcategories for a total of 14 subcategories. In reporting their results, Agarwal and Venkatesh focused on the categories and did *not* report results at the level of the subcategories. Although the categories do provide important information, they provide guidance at a higher level (e.g., ease of use) and it is difficult to translate such results directly into specific design guidelines for web sites. In contrast, an examination of the subcategories will shift the focus to more micro-level aspects of usability—for instance, goals rather than ease of use—that can be more directly related to *specific* design strategies such as those suggested by Nielsen (2000) and Keeker (1997). For example, per the results of Agarwal and Venkatesh, ease of use is weighted at 16.4 percent of the usability rating for customers in the airline industry. Let's say a site is rated very low on ease of use; then, design modifications to the site will need to focus on ease of use and will quite likely lead to broad or high level changes. In contrast, a focus on the three subcategories within ease of use (goals, structure, and feedback) may reveal the breakdown of weights to be 1 percent, 12 percent, and 3.4 percent respectively. Given the highest weighting being on structure, the design modifications could focus on micro-level ideas related to structure (see Keeker 1997). Therefore, we used both categories and subcategories in our work.

While the existing MUG categories and subcategories will apply to wireless sites, we expect to see some shifts in terms of the weights (i.e., different characteristics might gain prominence in this new context). Tied to the second objective, in the following subsection, we present the arguments related to the differences in the categories and subcategories that would be more important (i.e., weighted higher) in the context of web sites when compared to the categories and subcategories that would be more important (i.e., weighted higher) in the context of wireless sites. Tied to the third objective and building on the arguments presented next, in the subsequent section, we present a model of site use that includes the weighted rating (weight × rating) of various categories and subcategories as predictors of site use and web versus wireless as a moderator of key relationships.

Differences in Weights: Web Sites Versus Wireless Sites

In this subsection, we present the theoretical justification and hypotheses for differences in the weights placed on various categories in the evaluation of web sites versus wireless sites. Also, within each of the categories, we present the theoretical justification and hypotheses for differences in the weights on various subcategories in the evaluation of web sites versus wireless sites. Table 2 presents our hypotheses.

Table 2. Summary of Hypotheses

Category	Subcategory	Hypo.	Weight Differences—More important in:	Hypo.	Nature of Moderation:[a]
Content		H1(a)	Equal	H6(a)	No moderation
	Relevance	H1(b)	Wireless	H6(b)	Higher for wireless
	Media use	H1(c)	Web	H6(c)	Higher for web
	Depth/breadth	H1(d)	Web	H6(d)	Higher for web
	Timely/current information	H1(e)	Equal	H6(e)	No moderation
Ease of use		H2(a)	Wireless	H7(a)	Higher for wireless
	Structure	H2(b)	Wireless	H7(b)	Higher for wireless
	Goals	H2(c)	Equal	H7(c)	No moderation
	Feedback	H2(d)	Equal	H7(d)	No moderation
Made-for-the-medium		H3(a)	Wireless	H8(a)	Higher for wireless
	Personalization	H3(b)	Wireless	H8(b)	Higher for wireless
	Community	H3(c)	Web	H8(c)	Higher for web
	Refinement	H3(d)	Equal	H8(d)	No moderation
Promotion		H4	Web	H9	Higher for web[b]
Emotion		H5	Web; No subcategory hypotheses	H10	Higher for web[c] No subcategory hypotheses

[a]All effects from the MUG categories/subcategories are expected to be positive. The nature of the moderation alone is, therefore, shown in the column.
[b]Promotion is not expected to have a significant effect on wireless site use.
[c]Emotion is not expected to have a significant effect on wireless site use.

Content

In the context of web design, content has been identified as the most important characteristic contributing to the usability of web sites and as one of the key reasons that users return to web sites (Huzingh 2000; Lohse and Spiller 1998; Nielsen 1999; Ranganathan and Ganapathy 2002). Agarwal and Venkatesh found that content was weighted as the most important MUG category contributing to usability of web sites across multiple industries both among customers and investors. Content has been identified as the key factor in usability of wireless sites as well (Ballard and Miller 2001; Barnes et al. 2001; Ericsson Inc. 2002; Karkkainen and Laarni 2002; Palmer 2002). Given that individuals have to deal with devices with a small form factor (e.g., small screen, small keyboard), they will definitely place a high degree of emphasis on content in wireless sites. In sum, content will be an equally important category in contributing to usability of both web and wireless sites. Therefore, we hypothesize

H1(a): There will be no difference in the weight assigned to *content* across web sites and wireless sites.

The subcategories of content are relevance, media use, appropriate breadth or depth, and current or timely information. We argue that relevance will be more important in wireless sites, whereas media use and appropriate breadth/depth will be more important in web sites, and that there will be no difference in the importance of current or timely information across web and wireless sites. The studies identified earlier have examined several dimensions of content, for example, content organization, content adequacy, content layout (text versus graphics), use of links, etc. One common finding in studies that have examined web site usability is the importance of having *relevant content* readily accessible to the user, for example, the top of the information hierarchy (Czerwinski et al. 1998) and in the early pages (Nielsen 1999). Lohse and Spiller (1998) suggest that the availability of relevant content is one of the key reasons for a user's return to a web site. In the context of wireless sites, the ability to locate relevant content quickly has been identified as the single most important factor in wireless site use, including in white papers of organizations (see Ballard and Miller 2001; Ericsson Inc. 2002). Ramsay and Nielsen (2000) concluded that wireless sites offering services that focus on context-directed content are likely to be more successful than general wireless sites.

Finally, Barnes et al. (2001) found that questions relating to availability of relevant, accurate information were very important in wireless sites. This suggests that, as in web sites, the relevance subcategory will be a key determinant of usability in wireless sites. However, given that sites in a wireless context are expected to provide only a portion (subset) of the content and functionality of their web counterparts, we expect that the relevance subcategory will be more important in a wireless context. Therefore, we hypothesize

H1(b): The weight assigned to *relevance* will be higher in the context of wireless sites than it will be in the context of web sites.

The *media use* subcategory will be less important in evaluating wireless sites than it is in evaluating web sites. In traditional web sites and devices, where the multimedia support is extensive and high-speed connections are available, using media appropriately is key to a successful web presence. Although newer wireless, handheld devices support multimedia, including audio, video, and graphics, response latency and download speeds continue to be an issue (Ramsay and Nielsen 2000), thus resulting in very limited media use in wireless sites, which in turn is expected to render media use to be less important in evaluating wireless site usability. *Appropriate breadth and depth* is also likely to be more important in the context of web sites because of the comprehensive nature of information available on web sites. In contrast, on a wireless site, fewer killer functionalities will hold the key to success, thus rendering breadth to be less critical. Finally, *current or timely information* will not differentiate across web and wireless sites simply because in using sites in general, users expect up-to-date information. Therefore, we hypothesize

H1(c): The weight assigned to *media use* will be higher in the context of web sites than it will be in the context of wireless sites.

H1(d): The weight assigned to *appropriate breadth or depth* will be higher in the context of web sites than it will be in the context of wireless sites.

H1(e): There will be no difference in the weight assigned to *current or timely information* across web sites and wireless sites.

Ease of Use

Ease of use relates to the level of time and effort required to accomplish specific tasks (see Venkatesh 1999, 2000). Further, it is often a hurdle that must be passed and is more important to users in making decisions when the level of effort required is higher (Venkatesh 1999, 2000). Ease of use of a web site has been a subject of extensive study (Barnes and Vidgen 2002; Lederer et al. 1998; Lohse and Spiller 1999; Nielsen 1999). Prior research has examined the impact of the limitations of small screen size, lower screen resolution, and cumbersome input mechanisms (Buchanan et al. 2001; Duchnicky and Kolers 1983; Han and Kwahk 1994; Jones et al. 1999), and found that the time taken to complete specific tasks is greater, which in turn will result in a higher level of importance of ease of use in the context of wireless sites when compared to web sites. Duchnicky and Kolers (1983) evaluated the impact of screen size on readability and found that individuals could read larger screens 25 percent faster than screens that were one-third the size (see also Han and Kwahk 1994; Jones et al. 1999). Jones et al. (1999) found that users were 50 percent less effective in completing similar tasks when performing them on small screens compared to larger screens. Reduction in screen resolution can also erode task performance, which in turn affects device use (Buchanan et al. 2001). Given the pattern of findings with devices with larger screens, we expect the effect to be even more *pronounced* when the information access device is a wireless, handheld device that has a significantly smaller screen. Further, the limited input capabilities of devices such as cell phone keypads and "graffiti" pens are likely to elevate the importance of ease of use in the users' minds. Thus, it is likely that the ease of use category will be perceived to be more important for wireless sites than for web sites. Therefore, we hypothesize

H2(a): The weight assigned to *ease of use* will be higher in the context of wireless sites than it will be in the context of web sites.

The subcategories of ease of use are structure, goals, and feedback. We argue that structure will be more important in wireless sites, whereas goals and feedback will be equally important across web sites and wireless sites.

Of the various factors that contribute to ease of use of a web site, site structure and ease of navigation have been deemed important by several researchers. For example, Barnes et al. (2001) found "providing fast navigation to what I intend to find" to be one of the highest rated items in accessing wireless news sites (see also Barnes and Vidgen 2002; Nielsen 1999; Palmer 2002; Ranganathan and Ganapathy 2002). Consistent navigation links and a well-structured navigation hierarchy allow a user to move easily around a site. In the context of wireless sites, given the inherent limitations of handheld devices, we expect the structure subcategory to be more important than it would be in the context of web sites since a particularly well-structured wireless site can go a long way toward helping a user combat the limitations of the device. For example, Karkkainen and Laarni (2002) recommended that important information be made available high in the hierarchy in a wireless context since users are unlikely to navigate too deep. Kaikkonen and Ruto (2003) found that users of mobile phones preferred a flatter hierarchy to a deeper hierarchy, and tended to make extensive use of keyword search despite the input limitations of the devices. Thus, the structure subcategory is likely to be more important in a wireless context than it will be in a web context. Therefore, we hypothesize

H2(b): The weight assigned to *structure* will be higher in the context of wireless sites than it will be in the context of web sites.

Ease of use as a category was hypothesized to be more important in wireless sites (H2a). Based on the arguments above, we expect the increased importance of ease of use at the category level to be directed at structure subcategory. The subcategories of goals and feedback are expected to be comparably important across web sites and wireless sites and there are no substantive theoretical, empirical, or practical reasons to expect any differences. Therefore, we hypothesize

H2(c): There will be no difference in the weight assigned to *goals* across web sites and wireless sites.

H2(d): There will be no difference in the weight assigned to *feedback* across web sites and wireless sites.

Made-for-the-Medium

The made-for-the-medium category is tied to the user's ability to customize a site to suit his or her needs. The ability to customize web sites is a good design characteristic because it helps users save time and provides information that is of greatest interest to them, but even without it, users can access information with a few clicks on a traditional information access device. However, in the context of a wireless site, the ability to customize goes beyond a simple convenience and is essential when using a cumbersome device with a small keypad. Thus, we expect the made-for-the-medium category to be more important in a wireless context than in the web context. Therefore, we hypothesize

H3(a): The weight assigned to *made-for-the-medium* will be higher in the context of wireless sites than it will be in the context of web sites.

Of the various subcategories in this category (personalization, community and refinement) the one that has received the most attention in web site design is personalization. Personalization (via fewer key inputs and page accesses) can reduce the need for navigation in wireless sites (Karkkainen and Laarni 2002), and can attract return visitors to wireless sites (Serco 2000). For example, at AA.com's web site, users' preferences regarding preferred departure airports and destinations are captured during an initial visit. When a user revisits the site, any sale alerts to the users' preferred destinations are displayed. Another form of personalization is through the ability for a user to "jump to" relevant information on a web site. In wireless sites, personalization can reduce the need for navigation and combat the limitations of handheld devices such as limited screen size, navigational capabilities, and relatively slow connection speeds. We already see examples of such personalization in a wireless context. For example, cell phone operator portals, such as myVZW (Verizon Wireless) and Sprint PCS, allow users to customize the information they view on their handheld device. Given the inherent limitations of handheld devices, we expect the personalization subcategory will be more important in driving usability in a wireless context compared to a web context. Therefore, we hypothesize

H3(b): The weight assigned to *personalization* will be higher in the context of wireless sites than it will be in the context of web sites.

The notion of community in a web context can take various forms, such as participating in chat rooms or discussion forums, providing and viewing ratings for purchased articles, etc. Such facilities are intended to provide mechanisms that enhance the user experience. In a wireless context, the limitations of the device make it inconvenient to use community-oriented features since they typically require extensive navigation and typing. Finally, we have no *a priori* reasons to expect any differences between the web and wireless contexts for the refinement subcategory. Therefore, we hypothesize

H3(c): The weight assigned to *community* will be higher in the context of web sites than in the context of wireless sites.

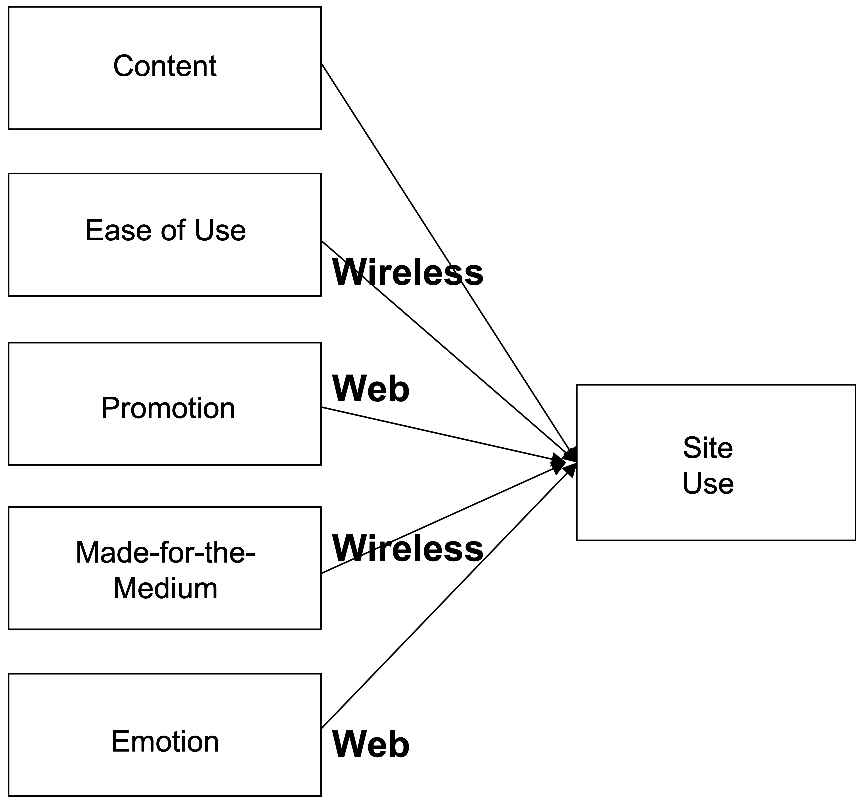
H3(d): There will be no difference in the weight assigned to *refinement* across web sites and wireless sites.

Promotion and Emotion

Given the added importance of ease of use and made-for-the-medium in a wireless context and the zero sum nature of the MUG instrument, we would expect the promotion and emotion categories to become less important in a wireless context. In today's environment of site use, users often first experience an organization's web site and then begin using the corresponding wireless site, which could render promotion to be a category of little importance in wireless site usability. In web sites, with one exception (i.e., customers of auto manufacturing web sites), Agarwal and Venkatesh found that the emotion category was weighted less than 10 points. We expect the emotion category to be weighted even less in the wireless context due to the business-oriented nature (rather than the entertainment or fun-oriented nature) of the sites examined in our research, and the functional and text-based nature of the wireless applications studied in our work. Also, given the low weight of the emotion category itself, we do not present separate hypotheses for subcategories as we expect subcategories in some cases to have almost no weight assigned in web and wireless contexts. Therefore, we hypothesize

H4: The weight assigned to *promotion* will be higher in the context of web sites than it will be in the context of wireless sites.

H5: The weight assigned to *emotion* will be higher in the context of web sites than it will be in the context of wireless sites.



Note: The words "Web" and "Wireless" indicate greater importance of weighted rating in the context of those types of sites.

Figure 1. Category-Level Predictors of Site Use

Models of Site Use

We present two models of site use. System use is a key dependent variable in IS research and is considered to be one of the important measures of system success (Delone and McLean 1992; Seddon 1997; Venkatesh et al. 2003). Research on technology adoption has extensively examined use as the primary dependent variable of interest (e.g., Davis et al. 1989; Taylor and Todd 1995; Venkatesh and Davis 2000). Specifically, we employ the various MUG categories and subcategories of usability as the predictors of site use. Also, web site versus wireless site is presented as a moderator of key relationships. Figures 1 and 2 present the proposed models of site use.

There are two important bases for the proposed models. First, HCI literature has demonstrated that usability, regardless of the specific conceptualization, drives use, be it for software or web sites (Agarwal and Venkatesh 2002; Klein 1998; Nielsen 2000; see also www.useit.com). Thus, the rich conceptualization of usability provided by the MUG conceptualization is expected to predict site—web and wireless—use. Second, the differences between web and wireless sites in terms of the importance of various categories and subcategories, discussed earlier, is at the heart of the moderating role of the effects of various categories and subcategories on site use. The arguments presented earlier focused on the relative importance of the weights of different categories and subcategories in the usability of web sites versus wireless sites. The weighted usability rating on any category or subcategory for a particular user for a particular site is the product of his/her assigned weight for that particular category or subcategory (the same for all sites in a specific industry for a specific user in a specific role) and his/her rating of that particular site on that category or subcategory. Therefore, our models of site use employ site use as the dependent variable and weight × rating of the various categories or subcategories as the predictors.

The arguments presented earlier tell us which categories and subcategories will be weighted higher in web sites versus wireless sites. When higher ratings occur in conjunction with higher weights, we can expect higher levels of use. It is possible for a category or subcategory to be weighted differently across web sites verus wireless sites. However, the implication of the difference on

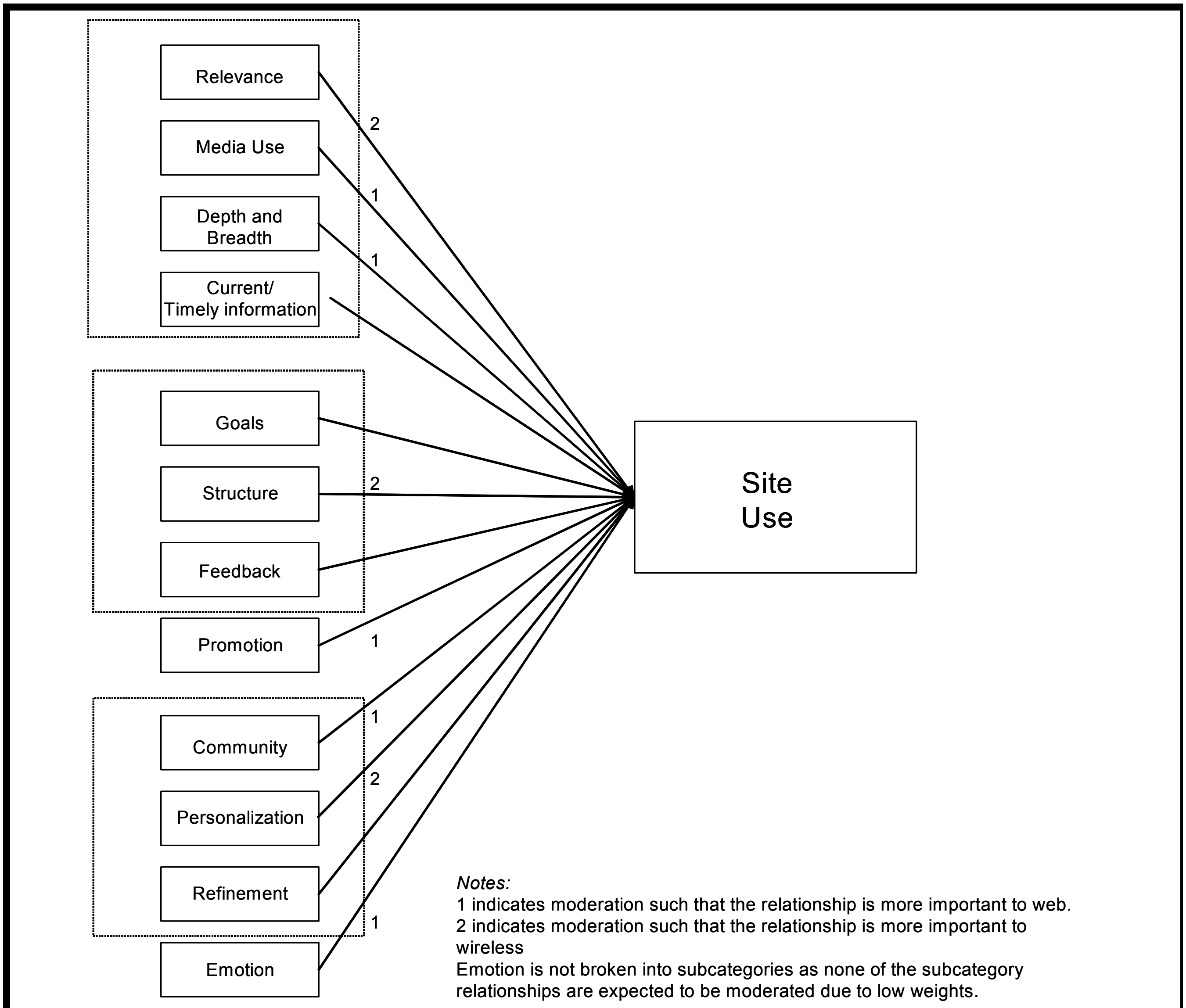


Figure 2. Subcategory-Level Predictors of Site Use

the dependent variable of interest (i.e., use) may be inconsequential because of the ratings, thus rending the effect on use to be non-significant. This suggests an important caveat associated with our proposed models. It is possible that we may not observe significant effects in the moderated models (e.g., web versus wireless) since weight differences could be washed out by ratings differences. This is alleviated in our models and the associated tests by the examination of separate direct effects only models, broken down by industry and information access device, which will help find the significant predictors of site use in each case.

Based on this, we present the following hypotheses:

H6(a): The weighted rating of *content* will positively influence web and wireless site use.

H6(b): The weighted rating of *relevance* will positively influence site use, with the effect being stronger on wireless site use.

H6(c): The weighted rating of *media use* will positively influence use, with the effect being stronger on web site use.

H6(d): The weighted rating of *appropriate breadth or depth* will positively influence use, with the effect being stronger on web site use.

H6(e): The weighted rating of *timely or current information* will positively influence web and wireless site use.

H7(a): The weighted rating of *ease of use* will positively influence site use, with the effect being stronger on wireless site use.

H7(b): The weighted rating of *structure* will positively influence site use, with the effect being stronger on wireless site use.

H7(c): The weighted rating of *goals* will positively influence web and wireless site use.

H7(d): The weighted rating of *feedback* will positively influence web and wireless site use.

H8(a): The weighted rating of *made-for-the-medium* will positively influence site use, with the effect being stronger on wireless site use.

H8(b): The weighted rating of *personalization* will positively influence site use, with the effect being stronger on wireless site use.

H8(c): The weighted rating of *community* will positively influence site use, with the effect being stronger on web site use.

H8(d): The weighted rating of *refinement* will positively influence web and wireless site use.

H9: The weighted rating of *promotion* will positively influence site use, with the effect being stronger on web site use and negligible on wireless site use.

H10: The weighted rating *emotion* will positively influence site use, with the effect being stronger on web site use and negligible on wireless site use.

In order to establish the new models of site use as a contribution, in addition to the richness of the set of categories and subcategories employed in MUG, it is important to establish its relative predictive validity vis-à-vis existing knowledge in the field. The most well-established and widely employed model of system use is the technology acceptance model (TAM; see Davis et al. 1989; for a review, see Venkatesh et al. 2003). TAM employs two key predictors: perceived usefulness and perceived ease of use (Davis et al. 1989). Perceived usefulness is defined as the extent to which a system increases one's productivity (Davis et al. 1989) and it has been applied in the context of personal use of technologies outside the workplace (e.g., Venkatesh and Brown 2001). Perceived ease of use is defined as the perceived extent of effort required to use the system (Davis et al. 1989; Venkatesh 2000). TAM has been applied extensively to study the adoption and use of various information technologies, including web sites (Venkatesh et al. 2003).

Since TAM is very parsimonious, it has been criticized for being unable to provide guidance to designers and managers to take remedial action or proactively manage the implementation. To establish the predictive validity of the models proposed, we compare them to TAM. Further, although the original theoretical development or subsequent extensions of TAM did not include web versus wireless as a moderator, we felt that arguments related to content and/or made-for-the-medium from MUG could apply to perceived usefulness from TAM, and ease of use from MUG could apply to perceived ease of use from TAM. Therefore, we included web versus wireless as a moderator in the appropriate empirical tests of TAM with site use as the dependent variable. The inclusion of such a moderated model also ensured that TAM was not being set up as a straw man. We expect that the model of site use based on MUG categories or subcategories as predictors given their broad range and breadth, particularly with web versus wireless as a moderator, will perform better than the context-independent TAM.

Method and Results

We conducted two studies to accomplish the three objectives outlined earlier. Specifically, the first study focuses on the first objective and aims to replicate and examine the generalizability of the MUG conceptualization, metric, and instrument from the United States to a new setting, Finland (see Lee and Baskerville 2003). Armed with the evidence of generalizability to Finland from study 1, the second study aims to examine the MUG categories or subcategories weight differences in both web and wireless contexts and to test the proposed models of site use (i.e., the second and third objectives).

Study 1

Participants and Sites

The settings for this study were universities in the United States and Finland. The population of interest in this study was Internet users. We identified 221 undergraduate students in the U.S. university by contacting instructors of senior-level electives in the business school. Students from a total of eight sections of different courses participated. No student participated more than once. Of these 221 participants, 201 provided usable responses. Six responses were incomplete and were dropped and 14 were eliminated due to misinterpretation of the questions in the survey instrument (discussed later). Of the 201 participants, 80 were women (39.8 percent). The average age was 21.02 with a standard deviation of 2.10. This study used one of the industries studied in Agarwal and Venkatesh (2002): the airline industry. The web sites studied were a subset of those studied in Agarwal and Venkatesh (2002): Northwest, Delta, and American Airlines.

The Finnish sample comprised 191 students.[4] The participants in Finland were also identified using a procedure similar to the one used in the United States. Students from a total of seven sections of different courses participated. No student participated more than once. Of the 191 participants, 169 provided completed responses, with 9 providing incomplete responses and 13 being eliminated from the data set because of misinterpretation of the questions in the instrument (discussed later). Of the 169 participants, 62 were women (36.6 percent). The average and standard deviation of age in the Finnish sample were 21.16 and 2.01 respectively. The sites tested among the Finnish participants were SAS and Finnair. These two airline web sites were chosen due to the expected familiarity with these two airlines and their web sites rather than the U.S. airlines and their web sites.

Procedure

The procedure used here was generally consistent with those used in Agarwal and Venkatesh. Participants were presented a description of the study and asked to take the role of a customer. The participants then provided their perceptions of the weights (i.e., relative importance) of the different categories and subcategories for web sites in the airline industry using a spreadsheet template that was provided to them. No kiosks were used in this study due to resource constraints and the logistical difficulties in conducting the study using kiosks in a university setting. One key distinction from Agarwal and Venkatesh was that in addition to providing the weights, participants were also asked to provide a brief justification for each of the category and subcategory weights that they assigned. We collected such data to ensure that the students in both the United States and Finland had the same frame of reference and interpreted the questions consistent with what was intended by Agarwal and Venkatesh. These justifications provide rich data about their understanding of each specific category or subcategory and their rationale for the weights assigned. Following this, the participants visited the sites using a browser. After browsing each site, the participants rated the site on each of the different MUG subcategories on a 10-point scale and provided the ratings on the same spreadsheet. Similar to providing a justification for each of the weights assigned, the participants provided a brief justification for the rating they assigned to each subcategory for each site (i.e., they explained what about a site prompted them to assign a certain rating for a particular subcategory for that site).

Results

Table 3 lists the means and standard deviations of the weights given by the participants to web sites in the airline industry in the United States and Finland. The weights assigned to various categories here were generally consistent with weights assigned by participants in the airline customer task of Agarwal and Venkatesh, thus establishing the face validity and initial credibility of our findings. We conducted MANOVA and ANOVAs with Scheffe's tests to examine the differences between the United States and Finland on the category weights, and found no significant differences. Table 4 presents the ratings[5] and the findings related to ratings' differences that were tested using MANOVA and ANOVAs with Scheffe's tests. We notice that the United States sites were rated somewhat comparably to what was found by Agarwal and Venkatesh for customers in the airline industry. The MUG metric discriminated across sites in the United States and Finland, lending credibility to our findings as they relate to Agarwal and Venkatesh.

While the empirical evidence discussed to this point is certainly important in terms of establishing the generalizability of the instrument, even more critical is ensuring that there was a common frame of reference across participants in both the United States and Finland. We carefully studied the justifications provided by the participants to ensure that the questions were interpreted correctly. Two evaluators—one of the researchers and a person not directly involved in the collecting of the data or the research—separately studied each respondent's justifications for each of the category weights, subcategory weights, and the subcategory ratings to ensure that it was consistent with the spirit of the construct. Although there are no statistical measures to test the common frame of reference, we computed each evaluator's assessment of the percent of violation (i.e., percentage of participants who misunderstood instructions and/or descriptions of the categories and subcategories). The number of misinterpretations (violations) were quite comparable—United States: 14 out of 201 (6.9 percent); and Finland: 13 out of 169 (7.7 percent). Although the violation rate in the Finnish sample was slightly higher, overall, it was still quite low. The inter-evaluator agreement reliability was computed to ensure that the interpretation of the correct and incorrect responses were consistent across the two evaluators—.97 and .98 in the United States and Finnish samples respectively. Based on these findings, there was strong evidence that the MUG instrument does generalize to Finland.

Study 2

Participants and Sites

The population of interest in this study was Internet and cell phone users. The sampling frame was visitors to a movie theater complex in downtown Helsinki, Finland, during a 1-week period. A total of

[4]The students were largely those who were pursuing a dual degree, that is, both undergraduate and graduate degrees in approximately 5 years. It is not possible to classify them as undergraduate or graduate students, but based on the number of credit hours completed, all of the students had completed less than 4 years of course work, thus being comparable to undergraduate students in the United States.

[5]It should be noted that in order to respect the anonymity of the sites, neither Agarwal and Venkatesh nor this research present the name of the site in conjunction with its usability rating.

Table 3. Study 1: Category Weights							
	USA		Finland		Significance of Difference	Agarwal and Venkatesh (2002): Customers in Airline Industry	
	M	SD	M	SD		M	SD
Content	32	5.75	35	6.16	NS	33	5.67
Ease of Use	12	4.88	10	4.01	NS	16	7.08
Promotion	9	4.01	7	3.12	NS	10	4.21
Made-for-the-medium	37	7.99	38	8.89	NS	32	8.08
Emotion	10	4.11	10	3.88	NS	8	3.77
Total	100		100			100	

Table 4. Study 1: Usability Ratings			
	Site 1	Site 2	Site 3
USA	7.10 (1.88)	6.45 (1.99)	7.89 (1.82)
Finland	7.33 (1.94)	6.12 (2.16)	N/A

Notes: 1. Ten-point scale used.
2. Means and standard deviations are shown; the latter is in parentheses.

Table 5. Study 2: Demographic Characteristics				
	Sites	N	Men/ Women	Age
Banking	Stock Alerts Yahoo! Finance	186	120/66	28.88 (7.11)
News	ABC News Guardian Unlimited	195	122/73	31.17 (7.91)
Shopping	Amazon Ebay	192	130/62	29.94 (6.55)
Tourism	CitiWiz Zagat Survey	193	125/68	33.12 (8.10)

Notes: 1. Means and standard deviations are shown for age; the latter is in parentheses.
2. The sites are listed alphabetically and do not represent the order in which they were entered in the data file.

812 individuals agreed to participate in the initial survey and 766 provided usable responses, for an effective response rate of 94.3 percent. The number of women in the sample was 287 (35.3 percent). The few participants who withdrew from the study after agreeing to participate indicated that the instrument was either too lengthy or they were unfamiliar with the site(s) and, therefore, not comfortable completing the survey. Further, we preferred that such participants withdraw from the study rather than provide inaccurate responses. The problems noted and actions taken by the researchers here are consistent with Agarwal and Venkatesh (2002). There were 13 unusable responses that were attributable to incomplete information. The demographic information of the participants broken down by the various industries is given in Table 5. In the follow-up survey conducted 3 months after the initial survey, 399 out of the 766 participants provided usable responses for a response rate of 52.09 percent. For the purpose of the model testing, these 399 respondents were used but for the analyses of the various mean differences between web and wireless sites, all 766 respondents were used, since it only required data from the initial survey.

Sites were chosen from four industries: banking, news, shopping, and tourism. The specific industries and sites chosen are shown in Table 5. The sites were chosen so that they had a web site accessible through a browser and a wireless web site that was accessible via a wireless (WAP-enabled) device. One of the key choices that needed to be made in the wireless context was regarding the use of a "live" mobile device versus an emulator. The use of emulators for development and testing, including usability testing, is a common practice in industry as evidenced by the fact that leading wireless, handheld device manufacturers such as Nokia, Motorola, and Palm provide emulators along with their application development toolkits. Given that the focus of our study is on the usability characteristics of wireless sites (e.g., their content, organization etc.), the use of an emulator eliminated extraneous factors such as network latency effects from potentially biasing the results. Hence, we used an emulator instead of a live wireless, handheld device in our kiosk. Specifically, we used the DeckIt emulator since it had a "skin" to emulate a generic wireless, handheld device not tied to any particular company's handset models.

Procedure

We conducted an initial survey to gather weights and ratings and a follow-up phone survey about 3 months after the initial survey to gather self-reported site use. The procedure used in the initial survey was highly consistent with Agarwal and Venkatesh. We also employed the mall-intercept method for participant solicitation. This study was conducted at a downtown movie theater complex in Helsinki, Finland. A promotion desk was staffed by up to four individuals, depending on time and day, with a "request to participate in a research study and get a free movie ticket and coupon for a popcorn and a drink." The incentive was valued at an approximate retail value of €15, with a validity of one year. Up to 10 kiosks were operational to allow the participants to browse the specific sites being studied and respond to the questionnaire. The use of 10 kiosks helped minimize participant wait time. In cases where there was a wait time, participants were given a numbered token. The numbers were called in sequence over a public address system. There were never more than 20 individuals waiting at any given time. In some cases, individuals received the token prior to their movie but came back and participated after their movie.

The Finnish participants, in general, were comfortable with conversational English and often visited English-language sites, especially those without Finnish counterparts. Therefore, in order to be consistent and control for the language, the participants were asked to visit English-language sites. About 10 participants who withdrew indicated that browsing English-language sites was too difficult for them. However, to ensure that participants understood the questions completely, we gave them a choice among English, Finnish, or Swedish versions of the questionnaire. Finnish and Swedish were used because they are both official languages of Finland. Professional translators, who routinely translate business documents and marketing brochures, translated the instructions and questions from English to Finnish and Swedish. Their translation procedure included a translation back to English by a different translator, and no discrepancies were found.

When a participant arrived at a kiosk, he or she was given a language choice for the questionnaire: English, Finnish, or Swedish. The participant was then assigned a particular industry chosen randomly by the computer. The participant was presented a description of the study and the customer role. We studied only one role (i.e., customer). We used a within-subjects design such that each participant provided his or her views on the web and wireless sites. Thus, in contrast to the design of Agarwal and Venkatesh, who compared across roles using a between-subjects design, we employed a more powerful within-subjects design for comparisons across information access devices. Using specific instructions and questions, each participant was prompted to provide his or her perceptions of the weights (i.e., relative importance) of the different categories and subcategories for web and wireless sites in that industry. Following this, the participants visited the web sites using a browser and wireless sites using the emulator. The order of presentation of the sites was randomized by the computer. Participants were given up to 5 minutes initially to browse each site. The computer then prompted the participant with a request to either continue browsing or rate the site; providing such additional time was consistent with Agarwal and Venkatesh, who argued that it would allow for more accurate ratings if the participant was not as familiar with the site. After browsing each site, the participant rated the site on each of the different MUG subcategories on a 10-point scale. Finally, demographic information was gathered. Based on system logs, it was determined that the average total time per participant spent browsing all sites was about 19 minutes and the average time spent filling out the survey was about 20 minutes. In addition to filling out questions regarding the various MUG categories and subcategories, participants also responded to four questions each on the perceived usefulness and perceived ease of use of specific sites. The items measuring these constructs were adapted from prior research that studied a variety of systems (for a review, see Venkatesh et al. 2003), including web sites (Koufaris 2002).

The follow-up survey was conducted about 3 months after the initial survey to gather information regarding participants' use of the sites (that they evaluated in the initial survey) in the past 3 months. The participants in the initial survey were called on their contact phone number, and they answered the use-related questions via phone. Of the 766 participants in the first wave of the study, 438 participants responded to the follow-up survey. However, 39 participants were excluded from the model tests because they did not have a phone that allowed access to the wireless sites, resulting in a final follow-up survey sample size of 399. Up to 10 callbacks were attempted to reach the participants and the follow-up calls were made over a 14-day period from 9 a.m. to 8 p.m. on weekdays and 11 a.m. to 8 p.m on Saturday and Sunday. The reasons for nonresponses were largely refusal to participate, no response at the called number, and disconnected phone numbers. Although we had no control over the nonresponse, an examination of the respondents in the follow-up with the nonrespondents did not reveal any statistically significant differences in terms of demographic characteristics and means or standard deviations of the importance and ratings. The specific questions used to gather information about use were consistent with prior research measuring use: number of times the site was visited in the past 3 months, amount of time spent on each visit, and assessment of intensity of use (Davis et al. 1989; Straub et al. 1995).

Means and standard deviations of the weights for the different categories and subcategories are shown in Tables 6 and 7 respectively. MANOVA confirmed differences in the category weights across information access devices. This was followed by ANOVAs and Scheffe's tests to understand specific differences. It should be noted that this is a within-subjects test since each participant provided weights governing web and wireless sites in the same industry. There were several interesting results, many of which highlight the advantages of examining subcategory-level weights.

1. Content was not significantly different across information access devices at the category level, thus confirming H1(a). However, it is worth noting the heightened importance of relevance in the wireless context in all four industries, thus supporting H1(b). While the weight placed on media use was somewhat low, it was more important in the web context in three out of the four industries, thus partially supporting H1(c). Depth and breadth was more important in the web context in all four industries, thus supporting H1(d). The weight placed on current and timely information was rather low and there was no difference in the weight assigned in three out of the four industries, thus partially supporting H1(e).

2. Ease of use was more important in a wireless context, thus supporting H2(a). While goals and feedback were equally important in the three out of the four industries in the web context, in the wireless context, the importance of structure was significantly higher than it was in the web context, thus lending support to H2(b, c, and d). Even where goals and structure differed across web versus wireless sites, the weight assigned was fairly low.

3. Made-for-the-medium was more important in the wireless context in three out of the four industries, thus partially supporting H3(a). Personalization was clearly the dominant characteristic in the wireless context and much more important in the wireless context than in the web context, thus supporting H3(b). Even in the one industry where there was no significant difference at the category level (i.e., news industry) between the web and wireless contexts, there was a significantly greater weight assigned to personalization in the wireless context. Community was more important in the web context than the wireless context, thus supporting H3(c). Refinement was comparably important in two of the four industries and there were small differences in the other two industries, thus partially supporting H3(d).

4. Promotion was minimally important in the web context and its importance declined even further in the wireless context, thus supporting H4.

5. The importance of emotion at the category level varied by industry in the web context; consistent with Agarwal and Venkatesh, emotion was more important in high involvement situations (i.e., here, shopping and tourism). However, with the minimalization of the importance of emotion in the wireless context, the subcategories conveyed little additional insight. This supported H5.

In addition to the interesting differences that emerged when comparing the weights assigned to web versus wireless contexts, there were significant differences in ratings when comparing the overall usability of an organization's web presence to its wireless presence. The ratings and the associated differences, assessed using MANOVA followed by ANOVAs with Scheffe's tests, are shown in Table 8. The ratings of web and wireless sites confirm the ability of the instrument to discriminate across sites within a given industry. For example, the ratings for the two wireless shopping sites were quite different. In comparing the ratings of sites within a given industry across information access devices, the wireless sites were rated lower in all cases. In fact, across all industries, the highest-rated wireless site (i.e., site 1 in the banking industry) was only slightly better than the lowest-rated web site (i.e., site 1 in the shopping industry).

The data were next analyzed to test the proposed model of site use. As noted earlier, in contrast to the previous analyses that focused on the differences in weights across web and wireless sites, this analysis focused on weight × rating of each category and subcategory and examined the prediction of *site use*. We used PLS-Graph, Version 2.91.03.04 to analyze the data as the dependent variable of site use. In order to test the model, we separated or pooled the data by indus-

Table 6. Study 2: Category Weights Across Industries and Information Access Devices

	Banking					News					Shopping					Tourism				
	Web		Wireless		Difference	Web		Wireless		Difference	Web		Wireless		Difference	Web		Wireless		Difference
	M	SD	M	SD		M	SD	M	SD		M	SD	M	SD		M	SD	M	SD	
Content	34	5.02	32	11.48	NS	30	5.99	31	10.78	NS	38	5.42	40	11.11	NS	33	5.36	34	7.98	NS
Ease of Use	16	2.90	27	8.12	***	15	2.10	26	7.62	***	13	1.85	23	6.69	***	15	3.16	29	6.68	***
Promotion	8	2.87	2	0.58	***	10	3.98	1	0.48	***	10	4.15	1	0.56	***	9	2.50	1	0.19	***
Made-for-the-Medium	34	5.25	38	12.25	*	38	4.65	40	13.25	NS	24	2.86	30	13.00	*	24	2.41	28	6.82	*
Emotion	8	2.05	1	0.29	**	7	1.15	2	0.39	**	15	2.59	6	0.89	**	19	2.20	8	1.32	***
Total	100		100			100		100			100		100			100		100		

Notes: 1. All the means have been rounded.
2. The "Difference" column indicates the significance of difference based on ANOVAs and Scheffe's tests.
3. NS: Non-significant; *p < .05; **p < .01; ***p < .001.

Table 7. Study 2: Category and Subcategory Weights Across Industries and Information Access Devices

	Banking					News					Shopping					Tourism				
	Web		Wireless		Difference	Web		Wireless		Difference	Web		Wireless		Difference	Web		Wireless		Difference
	M	SD	M	SD		M	SD	M	SD		M	SD	M	SD		M	SD	M	SD	
Content	**34**	**5.02**	**32**	**11.48**	**NS**	**30**	**5.99**	**31**	**10.78**	**NS**	**38**	**5.42**	**40**	**11.11**	**NS**	**33**	**5.36**	**34**	**7.98**	**NS**
Relevance	16	3.8	27	6.5	***	17	4.4	25	7.7	***	12	3.8	28	6.5	***	15	6.1	21	4.4	**
Media Use	4	0.6	1	0.3	*	4	0.6	1	0.3	*	4	0.8	3	0.6	NS	5	1.3	3	0.4	*
Depth/Breadth	6	1	1	0.4	**	5	1.1	1	0.3	*	15	4.9	2	0.3	***	5	1.8	2	0.3	*
Timely	8	1.7	3	0.6	*	4	0.5	4	1.1	NS	7	2.2	7	2.2	NS	8	2.4	7	2.1	NS
Ease of Use	**16**	**2.90**	**27**	**8.12**	*****	**15**	**2.10**	**26**	**7.62**	*****	**13**	**1.85**	**23**	**6.69**	*****	**15**	**3.16**	**29**	**6.68**	*****
Structure	8	2.1	20	4.1	***	6	0.8	18	4.8	***	5	0.7	15	5.6	***	7	1.6	17	5.3	***
Goals	6	1.4	3	0.6	*	7	0.9	7	2.1	NS	6	1	7	2.1	NS	7	1.8	9	2.8	NS
Feedback	3	0.5	4	0.8	NS	2	0.2	1	0.2	NS	2	0.4	1	0.4	NS	1	0.3	3	0.6	*
Promotion	**8**	**2.87**	**2**	**0.58**	*****	**10**	**3.98**	**1**	**0.48**	*****	**10**	**4.15**	**1**	**0.50**	*****	**9**	**2.50**	**1**	**0.19**	*****
Made-for-the-Medium	**34**	**5.25**	**38**	**12.25**	***	**38**	**4.65**	**40**	**13.25**	**NS**	**24**	**2.86**	**30**	**13.00**	***	**24**	**2.41**	**28**	**6.82**	***
Personalization	14	4.7	29	5.1	**	17	4.5	31	6.1	***	8	2.8	28	4.7	***	9	2.5	17	4.7	**
Community	6	1.4	1	0.5	**	11	2.9	1	0.3	***	8	3	0	0.1	***	7	2.1	2	0.3	**
Refinement	14	4	8	1.8	*	10	2.7	8	1.3	NS	8	2.7	2	0.4	**	8	2.2	9	2.5	NS
Emotion	**8**	**2.05**	**1**	**0.29**	****	**7**	**1.15**	**2**	**0.39**	****	**15**	**2.59**	**6**	**0.89**	****	**19**	**2.20**	**8**	**1.32**	*****
Challenge	0	0.1	0	0.1	NS	0	0.1	0	0.1	NS	0	0.1	0	0.1	NS	3	0.5	1	0.3	*
Plot	0	0.2	0	0.1	NS	2	0.4	~	0.1	*	4	0.2	2	0.3	*	5	0.7	2	0.5	*
Character Strength	4	0.5	~	0.2	*	3	0.6	1	0.3	NS	4	0.3	2	0.3	*	8	1.4	4	0.8	*
Pace	4	0.6	~	0.2	*	2	0.2	~	0.1	*	7	1.4	2	0.3	**	3	0.3	1	0.3	*
Total	100		100			100		100			100		100			100		100		

Notes: 1. All the means have been rounded up. The ~ sign indicates a factional number that approximates to zero when rounded up.
2. The "Difference" column indicates the significance of difference based on ANOVAs and Scheffe's tests.
3. NS: Non-significant; *p < .05; **p < .01; ***p < .001.

Table 8. Study 2: Usability Ratings for Web and Wireless Sites

Industry		Site 1	Site 2	Differences Across Sites
Banking	Web site	7.12 (2.12)	6.42 (2.09)	*
	Wireless site	6.10 (2.07)	5.53 (2.01)	*
	Differences across information access devices	**	**	
News	Web site	7.20 (1.05)	7.10 (1.02)	NS
	Wireless site	4.23 (1.03)	5.10 (1.02)	**
	Differences across information access devices	***	***	
Shopping	Web site	6.04 (1.14)	7.81 (1.09)	***
	Wireless site	3.27 (1.13)	7.16 (1.31)	***
	Differences across information access devices	***	*	
Tourism	Web site	7.09 (1.41)	7.13 (1.12)	NS
	Wireless site	5.16 (1.07)	4.01 (1.09)	**
	Differences across information access devices	***	***	

Notes: 1. Means and standard deviations are shown; the latter is in parentheses.
2. The difference column indicates the significance of the difference based on a Scheffe's test.
3. *p < .05; **p < .01; ***p < .001; NS indicates nonsignificant difference.
4. Ten-point scale used.

try and information access device as needed. In the fully pooled analysis, this resulted in a sample of 399 × 4 = 1,596. Each respondent thus contributed four data points for this analysis as they provided distinct data records for each site. The dependent variable was self-reported use that comprised the following formative indicators: frequency, duration, and intensity. The independent variables were the various weighted category ratings for each site. This was calculated as a product of the weight assigned to each category for the particular industry multiplied by the rating assigned to the particular site on that category. This is consistent with the approach outlined in Agarwal and Venkatesh to calculate the usability rating for each category. The web versus wireless site was coded using a dummy variable, SITE-TYPE, which was coded as 0 for a web site and 1 for a wireless site. All latent variables, with the exception of use, were represented by a single indicator variable due to the nature of the MUG metric and the instrument. Interaction terms were created consistent with Chin et al. (2003).[6] In addition to the main effects of the various weighted category ratings, the dummy variable, SITE-TYPE, was included. Also, for those relationships marked as moderated in Figure 1, interaction terms were created as a product of the indicator for the weighted category rating and SITE-TYPE. Similar to the approach described above, we set up structural models at the level of the subcategories, including direct effects only models and moderated models as shown in Figure 2, broken down by industry and pooled across industries.

Table 9 presents the results of the model testing at the category level broken down by industry with separate model tests for web sites, separate model tests for wireless sites, and pooled across web and wireless sites (including web versus wireless as a moderator). In the direct effects only models, content, ease of use, and made-for-the-medium were significant determinants of web sites and wireless sites. Emotion was significant in shopping and tourism web sites alone. In a test of the pooled models, in all four industries, SITE-TYPE did not moderate the effect of content on site use. Ease of use had a stronger effect in wireless sites in all four industries. Simi-

[6]Although PLS produces a measurement model, in our analyses, the measurement models were not very pertinent given that most latent variables were represented by a single indicator variable resulting in a loading of 1, and the AVEs and ICRs were 1. The weights of the three formative indicators for site use were .49, .52, and .54 respectively in the fully pooled model; in other models also, the weights were in the range of .40 to .59. Since the indicators for site use were formative, the weights are used to assess the relative contribution of each indicator in the formation of the construct, rather than as a way to assess validity of the measure.

Table 9. Results of Model Testing at the Category Level

	Banking			News			Shopping			Tourism			Pooled Across Industries		
	Web	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**
R^2	**.45**	**.59**	**.74**	**.46**	**.60**	**.72**	**.50**	**.60**	**.73**	**.46**	**.58**	**.73**	**.46**	**.59**	**.72**
Adjusted-R^2	**.41**	**.53**	**.66**	**.41**	**.54**	**.65**	**.44**	**.55**	**.65**	**.42**	**.53**	**.67**	**.41**	**.53**	**.65**
Content	.28***	.27***	.27***	.25***	.25***	.25***	.33***	.34***	.34***	.28***	.28***	.28***	.27***	.29***	.28***
Ease of use	.15**	.25***	.08	.15**	.23***	.04	.14*	.21***	.04	.14*	.25***	.04	.15*	.24***	.04
Promotion	.04	.02	.10	.07	.02	.07	.07	.00	.02	.03	.04	.02	.03	.04	.02
Made-for-the-medium (MFTM)	.21***	.28***	.10	.22***	.30***	.12*	.21***	.28***	.05	.20***	.29***	.07	.20***	.29***	.04
Emotion	.09	.00	.08	.06	.04	.02	.15*	.04	.04	.18**	.07	.04	.11*	.02	.02
SITE-TYPE (0: Web; 1: Wireless)			.14*			.06			.04			.02			.05
Content × SITE-TYPE			NA			NA			NA			NA			NA
Ease of use × SITE-TYPE			.31***			.28***			.25***			.28***			.29***
Promotion × SITE-TYPE			.02			.04			.04			.01			.05
MFTM × SITE-TYPE			.16**			.15*			.18**			.20***			.18**
Emotion × SITE-TYPE			.01			.01			-.16**			-.15*			-.10*

Note: $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$; NA indicates not applicable

Table 10. Results of Model Testing at the Subcategory Level

	Banking			News			Shopping			Tourism			Pooled Across Industries		
	Web	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**
R^2	**.59**	**.55**	**.75**	**.58**	**.59**	**.76**	**.58**	**.59**	**.77**	**.65**	**.66**	**.75**	**.60**	**.60**	**.76**
Adjusted-R^2	**.55**	**.53**	**.67**	**.53**	**.56**	**.68**	**.53**	**.56**	**.68**	**.56**	**.55**	**.67**	**.54**	**.57**	**.69**
Relevance	.17**	.25***	NS	.17**	.27***	NS	.13*	.28***	NS	.20***	.27***	NS	.17**	.27***	NS
Media use	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Depth and breadth	NS	NS	NS	NS	NS	NS	.15*	NS	NS	NS	NS	NS	NS	NS	NS
Current/timely info	NS	NS	NS	NS	NS	NS	NS	NS	NS	.13*	.13*	.13*	NS	NS	NS
Structure	.13*	.22***	NS	.12*	.20***	NS	NS	.25***	NS	.13*	.25***	NS	.13*	.24***	NS
Goals	NS	NS	NS	.11*	NS	NS	NS	NS	NS	.13*	.14*	.13*	NS	NS	NS
Feedback	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Promotion	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Community	.13*	NS	NS	.13*	NS	NS	.13*	NS	NS	.14*	NS	NS	.13*	NS	NS
Personalization	.14*	.22***	NS	.16**	.30***	NS	.14*	.33***	NS	.14*	.22***	NS	.15*	.28***	NS
Refinement	.14*	NS	NS	.13*	NS	NS	.13*	NS	NS	.13*	.13*	.13*	.14*	NS	NS
Challenge	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Plot	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Character strength	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Pace	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
SITE-TYPE (0: Web; 1: Wireless)			NS			NS			NS			NS			NS
Relevance × SITE-TYPE			.28***			.33***			.30***			.29***			.29***
Media use × SITE-TYPE			NS			NS			NS			NS			NS
Depth and breadth × SITE-TYPE			NS			NS			-.14*			NS			NS
Current/timely info × SITE-TYPE			NS			NS			NS			NS			NS
Structure × SITE-TYPE			.25***			.28***			.30***			.26***			.30***
Goals × SITE-TYPE			NS			-.11*			NS			NS			NS
Feedback × SITE-TYPE			NS			NS			NS			NS			NS
Promotion × SITE-TYPE			NS			NS			NS			NS			NS
Community × SITE-TYPE			-.17**			-.16**			-.16**			-.16*			-.16**
Personalization × SITE-TYPE			.26***			.29***			.28***			.25***			.30***
Refinement × SITE-TYPE			-.16**			-.13*			-.14*			NS			-.15*
Challenge × SITE-TYPE			NS			NS			NS			NS			NS
Plot × SITE-TYPE			NS			NS			NS			NS			NS
Character strength × SITE-TYPE			NS			NS			NS			NS			NS
Pace × SITE-TYPE			NS			NS			NS			NS			NS

Note: p < .05; **p < .01; ***p < .001; NS indicates nonsignificant larly, made-for-the-medium had a stronger effect in all four industries. Finally, promotion did not play a role, direct or moderated, in any industry. Emotion, however, was more important in web sites in two of the four industries, and completely nonsignificant in the other two. This pattern of results supports H6(a), H7(a), H8(a), and H10, and partially supports H9.

The data were then pooled across industries. The results of the model testing for web sites separately (direct effects only), wireless sites separately (direct effects only), and pooled (web and wireless) are also shown in Table 9 (in the right-most three columns). The results from the pooled analysis further support the patterns seen in the industry-by-industry tests. Content was comparably important across web and wireless sites, thus supporting H6(a). Ease of use and made-for-the-medium were more important in wireless sites, thus supporting H7(a) and H8(a). Promotion did not predict site use, thus not supporting H9 in the pooled test. Emotion was more important in web sites, thus supporting H10.

Table 10 presents the results of the subcategory-level analyses broken down by industry and reported separately for web sites, separately for wireless sites, and pooled across web and wireless sites within each industry (including web versus wireless as a moderator). The findings related to the critical subcategories were very consistent across industries. The important drivers of site use across both web and wireless sites were relevance, structure, and personalization. In addition, in the context of web sites, community and refinement were consistent drivers in all industries. While there were more drivers in the case of web sites, in wireless sites, the three consistently significant drivers (relevance, structure, and personalization) had stronger effects on wireless site use than those drivers did in determining web site use. The pooled analysis confirmed this as the effects of relevance, structure, and media use on site use were moderated by SITE-TYPE such that they were more important among wireless sites. The effects of community and refinement on site use were moderated such that they were more important among web sites. Overall, this pattern confirms the key arguments and hypotheses related to the subcategory-level hypotheses in H6, H7, and H8. The data were then pooled across industries. The results of the model testing for web sites separately (direct effects only), wireless sites separately (direct effects only), and pooled (web and wireless) are also shown in Table 10 (the right-most three columns). The results from the pooled analysis further support the patterns seen in the industry-by-industry tests. Once again, relevance, structure, and personalization were more important drivers of site use among wireless sites, while community and refinement were more important drivers of site use among web sites. The pooled analysis indicated that the interaction terms related to these five subcategories were the only determinants of site use. This analysis also supported the general arguments and subcategory-level hypotheses in H6, H7, and H8. However, it should be noted that not all of the predicted moderators were significant. As we noted earlier, some of these nonsignificant effects, especially on low-weighted subcategories, may be due to the differences in ratings' washing out any weight differences and, consequently, having a null effect on web versus wireless site use.

Finally, to benchmark our models, we tested TAM in the various industries and also pooled across industries. We tested both the baseline TAM per Davis et al. (1989) and also the moderated model including web versus wireless as a moderator. The interaction terms were created as suggested by Chin et al. at the indicator level. The various measurement models showed reliability and validity for the perceived usefulness and perceived ease of use scales consistent with much prior research on TAM (for a review, see Venkatesh et al. 2003) and, therefore, these results are not shown here in the interest of space. Table 11 presents the results of the model testing. It is not surprising that the basic TAM is supported in all industries in web and wireless contexts given its extensive previous validation in a variety of contexts. Consistent with what was observed in MUG model tests related to content and ease of use respectively, web versus wireless was not a moderator of the perceived usefulness-use relationship but was a moderator of the perceived ease of use-use relationship. However, the key pattern that lends support to the new model is that richness and amount of variance was greater for the new model of site use based on MUG (i.e., about 70 percent) when compared to any of the TAM tests (i.e., about 50 percent), thus underscoring that the new model provides theoretical understanding and empirical superiority in predicting use of web and wireless sites.

Discussion

The current work employed the conceptualization, metric, and instrument based on Microsoft Usability Guidelines (MUG) to further our understanding of the differences in usability between web and wireless sites and their ties to site use. Two studies, including one with longitudinal data collection, were conducted to accomplish the various research objectives. Based on our theoretical development and empirical evidence, we accomplished three objectives: established the generalizability of MUG to a new setting, understood the differences between web and wireless usability, and developed and tested a model of site use based on MUG.

Theoretical Contributions, Implications, and Future Research Directions

The theoretical contributions of this work are significant. Following the recommendations of Lee and Baskerville (2003), our work established generalizability of the MUG metric and instrument by examining its applicability in settings different from the work that introduced the MUG conceptualization, metric, and instrument (i.e., Agarwal and Venkatesh 2002). Our first study demonstrated generalizability from the United States to Finland. Specifically, based on a careful assessment of the justifications provided for

Table 11. Results of TAM Testing

	Banking			News			Shopping			Tourism			Pooled Across Industries		
	Web	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**	**Web**	**W'less**	**Pooled**
R^2	**.34**	**.40**	**.50**	**.36**	**.42**	**.53**	**.36**	**.42**	**.52**	**.39**	**.47**	**.56**	**.37**	**.44**	**.52**
Adjusted-R^2	**.33**	**.39**	**.48**	**.35**	**.40**	**.51**	**.35**	**.40**	**.49**	**.38**	**.46**	**.54**	**.36**	**.43**	**.49**
Perceived usefulness (PU)	.34***	.36***	.35***	.33***	.36***	.35***	.35***	.32***	.34***	.38***	.35***	.35***	.34***	.35***	.34***
Perceived ease of use (PEOU)	.20**	.28***	.04	.23***	.31***	.07	.22***	.32***	.10	.18*	.38***	.02	.20**	.35***	.10
SITE-TYPE (0: Web; 1: Wireless)			.10			.02			.08			.06			.03
PU × SITE-TYPE			.05			.03			.04			.05			.00
PEOU × SITE-TYPE			.36***			.40***			.44***			.41***			.35***

Notes: *p < .05; **p < .01; ***p < .001

weights and ratings assigned by each individual, we conclude that the instrument can be applied outside the United States, particularly in Finland, because of a common frame of reference among participants in both settings. It is, of course, important to recognize that the complete achievement of the implicit ideas underlying the first objective (i.e., complete generalizability) will take years of concerted work by many researchers. Along the way, it is quite likely that modifications to the MUG conceptualization, metric, and instrument will be necessary to be sensitive to cultural contexts, meanings, and symbolisms. The generalizability to Finland, while important, should be viewed in light of the fact that Finland, like the United States, comprises a highly literate population with a fairly high degree of sophistication. As the generalizability to other countries such as China and India are examined, additional categories and subcategories may also need to be considered to fully understand what usability means to people in those countries.

The second study helped us generalize the MUG conceptualization, metric, and instrument to different industries (i.e., banking, news, shopping, and tourism) from the ones studied in Agarwal and Venkatesh (2002) in the new country (i.e., Finland). Further, this study also helped establish the use of MUG to a different information access device (i.e., wireless, handheld device). Finally, the theoretical model of site use that employed MUG categories and subcategories as predictors with web versus wireless as a moderator received strong support in all industries studied, and provided greater richness than the most widely used traditional model of use (i.e., TAM) and explained greater variance than TAM in all industries. This pattern is supportive of the view of Benbasat and Zmud (2003) that the *information technology artifact* needs to be more central to theory development in information systems, as the more context-specific model of site use proposed here was superior to the context-independent TAM, both in terms of richness of guidance to site designers and variance explained. In recent years, TAM has become somewhat of a gold standard for understanding individual reactions to technology and use behavior given its combination of parsimony, reliable and valid scales, and generalizability to a range of contexts and technologies. While TAM's strengths are indisputable, this paper suggests that a somewhat more context-specific theory provides greater explanatory power and more actionable guidance to practitioners. Although this work is one of the first investigations comparing TAM, a model based on general cognitive beliefs about technology, with a model based on perceptions of design attributes, the evidence here supports the latter model. The findings here start to expose some of the boundary conditions of TAM, identify meaningful directions for further inquiry in technology adoption and use, and address the need for guidance for IS designers (a standing criticism of TAM).

In contrast to Agarwal and Venkatesh, who reported only category-level results, our work focuses on both category and subcategory levels. This raises the debate of richness versus parsimony in the context of the MUG metric. The debate of richness versus parsimony has existed in the reference disciplines for decades from the perspective of theory development and analytical techniques (Battig 1962; Chaiken et al. 1999; Ressler 1963; Sivo and Willson 1998). IS researchers have also debated this issue; some have weighed in favor of parsimony (Davis et al. 1989) and others in favor of richness (e.g., Plouffe et al. 2001; Taylor and Todd 1995). The data from our second study suggests that the richness of the information garnered by focusing on the subcategories in MUG outweighs the parsimony that comes with a focus on categories alone. The support for our models, when examined in light of the support for TAM, underscores the richness and depth of our model.

It is also important to note that while the Microsoft Usability Guidelines (Keeker 1997) were developed based on the results of usability testing, there is no evidence to indicate that the categories and subcategories identified there are comprehensive. Human-computer interaction is rich with several guidelines and suggestions for effective site design. Future research should examine the possibility of adding or removing more categories and subcategories to the guidelines based on both theoretical and practical considerations. Furthermore, additional contingencies as they relate to web versus wireless sites should be identified, particularly as they relate to new categories and subcategories that are added.

Our model of site use was very predictive of site use and provided rich insights but comes from a single paradigm of work (i.e., human-computer interaction). Recent work in IS integrated eight models of adoption and use with roots in IS, psychology, and sociology, and explained over 70 percent of the variance in key outcomes (see Venkatesh et al. 2003). Similarly, there is work explaining adoption and use in non-organizational contexts drawing primarily from research in IS, psychology, sociology, and consumer behavior (e.g., Kraut et al. 1999; A. Venkatesh and Vitalari 1992; Venkatesh and Brown 2001). Such competing views should be compared with the MUG-based model of site use. Further, the different models should be integrated as it is likely that predictors from different perspectives would contribute to a richer and more holistic model of site use. An integrated model would provide a better and more complete nomological network (see Straub 1989; Venkatesh et al. 2003; Venkatraman and Grant 1986) and will allow for predictors from competing models to be simultaneously examined to determine the most important predictors. While integrative models are one general direction for future work, micro-level research—typically, experiments—should examine various design strategies and their effect on user ratings and outcomes such as site use. Finally, while site use is an important measure of system success (Delone and McLean 1992; Seddon 1997), it is important to examine other outcomes of broader organizational interest such as site loyalty, switching costs, information satisfaction, and online sales. Research on the extent to which the usability categories and subcategories identified here contribute to such outcomes will be important work at the intersection of IS and consumer behavior. Consumer behavior literature has a rich set of predictors of many of these outcomes and the predictors here will quite likely be complementary and provide a richer understanding of the core underlying phenomena.

One final direction for future research is to further our understanding among various user groups. Agarwal and Venkatesh studied two different user groups (customers and investors) and found many intergroup differences in what was important in assessing usability. We focused only on customers and future work is essential to understand the applicability of MUG and the model proposed here to a broader set of user groups (e.g., suppliers and employees) as well as segmentation based on other demographic characteristics. Not only is it likely that the factors that are important will be different (see Agarwal and Venkatesh 2002) but also, it is quite likely that the set of categories and, of course, subcategories will need to be expanded, thus providing the opportunity for further development of the metric and instrument. It also presents the opportunity for researchers to focus on expanding the research model presented here.

Practical Implications

The results from the research are of value to practitioners as well. The MUG conceptualization, metric, and instrument is one of the first holistic usability evaluation methods available that included a detailed application procedure tailored to end-users (Agarwal and Venkatesh 2002). The findings from the current work should assure practitioners that the instrument can be applied to a range of industries, can be applied successfully to samples outside the United States, and to wireless sites. The application of the instrument, particularly to the wireless context, strengthens its validity in today's world of technological convergence where it appears that the wireless, handheld device may very well be the dominant information access platform in the not-to-distant future.

The strong contrast in the weights between web and wireless contexts for any given industry has implications for wireless site design. Further, the significant differences in ratings for any given industry's web sites compared to their wireless sites suggests that much work needs to be done in order to design effective and usable wireless sites. The current research also demonstrated that a dominating web presence may not necessarily translate into a dominant wireless presence. Organizations will be well-served to not be complacent with their wireless site design given its growing importance. Our results provide specific areas of focus that could be targets for redesign efforts, particularly for the industries and sites that we studied. With the procedures and validity further reinforced by this work, practitioners in other industries can conduct usability studies using this instrument in order to understand the overall weighting scheme used by customers in their industry vertical. Further, by benchmarking their site against competitors' sites, they can identify weak areas that can be a focus of their wireless site redesign efforts.

Our results provide solid empirical evidence that wireless content providers need to pay particular attention to assessing what *core content* is provided through their site since it is a key predictor of use. In a wireless context, it is quite likely that customers will perceive only a subset of the available services to be useful. At the extreme, a few key services might become the *required killer functionality* for a given industry vertical (i.e., the key reasons why a customer accesses a particular site using their wireless information access device). For example, in the context of a personal banking web site accessed using a cell phone, it is possible that the *killer functionality* would be the ability to check account balances. Other functionality such as transferring balances or applying for loans might seldom be accessed using a handheld device.

The *structure* of a wireless site also emerged as a key aspect of wireless usability. Hence, wireless site designers should follow guidelines for structuring wireless sites such as always having a back button and a home button, not requiring users to scroll in more than one direction, creating content that is specifically meant for the wireless context, using advances in technologies such as XHTML and style sheets to create "generic" wireless content that can be customized to the recipient device and encouraging the use of gadgets such as attached keyboards or track wheels. It goes without saying that all of the above guidelines only bear fruit when the content itself has been pruned such that only *relevant content* is being displayed. In addition, designers should also consider findings from research that solves some issues that stem from device limitations, for example, the use of rapid serial visual presentation (RSVP) to overcome the screen size limitation (Castelhano and Muter 2001).

The results also clearly highlight the critical role that *personalization* technology can play in a wireless context. We identified some mechanisms for personalization already in use in a wireless context. Perhaps the most intriguing emerging form of personalization is one that combines user identification data (e.g., by using "key information" about their devices), with information about their current location. For example, a wireless shopping site can present users with focused content based on prior knowledge of their preferences and knowledge of their current location (e.g., proximity to a retail clothing store that has a preferred item on sale). Using this strategy, the wireless site can provide *relevant, timely content* through the use of *personalization* techniques, thus simultaneously leveraging more than one key driver of wireless site usability. Another form of personalization can lead users directly (based on specified preferences) to different parts of a wireless site directly, thus saving them the effort of navigating through multiple pages. For example, a user can be led directly to a sports ticker page every time they click on a sports portal instead of the generic sports news, if that is what they prefer. In addition, the navigational links from a given page can also be customized to the user (e.g., present information about their favorite sports team, thus reducing the navigational effort required).

Finally, according to our results, the *community-oriented characteristics* are more important in a web context when compared to a wireless context. One possible explanation for this is that in a wireless context the group-oriented features are captured through other channels such as short message service (SMS). Further, as evidenced by our earlier results on content, users in a wireless context tend to have more of a "purpose" when visiting sites. Thus, traditional community-oriented features such as community ratings for products or chat forums for information exchange are not likely to be useful in a wireless context. Hence, wireless site designers should not waste valuable screen real estate providing such information. However, more context-aware community-oriented features might be useful in a wireless context, for example one that notifies users of the presence of other users, say from a "buddy" list, online or within a certain geographical area, such as Carnegie Mellon University's MyCampus (http://www-2.cs.cmu.edu/~sadeh/mycampus.htm).

Limitations and Additional Future Research Directions

One of the limitations of this work stems from the use of an emulator instead of a "live" cell phone to access the wireless sites. This choice was made primarily to be able to manage the logistics of getting each user to access both a web and wireless site (in a random order) within the constraints of the space allocated to us at the movie theater complex. Other constraints that played a role in our decision were the need for us to able to monitor the users in a kiosk setting and ensuring that users were able to complete the task within a reasonable amount of time. In addition, as noted earlier, we wanted to eliminate any effects due to network performance issues. We would like to note that issues surrounding the use of the emulator are partially mitigated in the sense that the use of the emulator should not impact the weights assigned to categories or subcategories (for web or wireless sites) since the participants provided these weights before visiting the sites using the emulator. It is, however, likely that the use of an emulator keypad on the computer (instead of an actual cell phone keypad) for input might have had an effect on the ratings. Future research can get a more accurate assessment of wireless site ratings by using live wireless, handheld devices. This is important since this research likely presents a rosier picture of wireless site ratings because of the use of kiosks and emulators whereas tests with the real devices might reveal that the state-of-the-art is actually worse. We also believe that, given the advancements made in the current generation of wireless networks in terms of their increased speed (2.5G networks are widely available and even 3G networks are available) and the reliability of data- oriented services, it might be time to extend and replicate Ramsay and Nielsen's (2000) study to assess the nature of the wireless experience in the context of today's wireless networks.

Our choice of wireless sites was based on the popularity of the site (as evidenced by it being listed on sites such as CNET.com or Yahoo!) as well as the availability of a reasonable amount of content without the need for a login. The latter constraint meant that we were not able to use online banking sites in our study. Instead, we were limited to using sites that provided access to free financial information. Another limitation of our study is that we exposed users to a single generic cell phone model in the emulator. However, in today's environment, it is possible to access wireless sites from several different types of devices (e.g., PDAs, smartphones, etc.). While advances in technologies such as the .NET compact framework (http://www.microsoft.com/net/) and J2ME (http://java.sun.com/j2me) make it easy for developers to render applications across different types of wireless, handheld devices, assessing the usability of their wireless sites on each of these devices is still important. Thus, further studies focusing on the usability of wireless sites should also try to use different types of devices and, possibly, various models within each type of device. Such specific research studies will help further understand specific design practices and their impact on various outcomes.

Earlier, we discussed some of the limitations of our work as it relates to generalizability to new countries; here, we further expound on that point. For practical reasons, our work examined the applicability of the MUG instrument in one new country (i.e., Finland) but, as with all cross-cultural research, future research should examine the generalizability to other countries. With the extensive diffusion of cell phones and various mobile services in Asia, research examining the generalizability of our findings and/or explicating differences in the major Asian markets (e.g., China, Hong Kong, India, Singapore, and South Korea) are crucial since the Asian markets have become a dominating presence in the wireless space. Such work will not only serve an important scientific purpose when supported by theoretical arguments regarding expected cross-cultural similarities and differences (Lee and Baskerville 2003), but also provide valuable insights for practitioners to better understand today's global village.

Conclusions

We used the MUG metric and instrument to understand wireless site usability as well as to understand the differences between web site and wireless site use. Since we conducted our study in Finland, we first established the generalizability of the metric and instrument by examining its validity in Finland, a different setting from the one in which the MUG metric and instrument were originally developed and validated. Armed with the evidence of generalizability, we examined the differences between the weights placed by users in assessing web site versus wireless site usability. The key findings were the higher weights assigned to the categories of ease of use and made-for-the-medium in wireless contexts, with a specific emphasis on the subcategories of relevance, structure, and personalization. Finally, the models of site use using MUG categories or subcategories as predictors explained about 70 percent of the variance in site use when compared to TAM's approximately 50 percent. In sum, the results here provided strong support for the use of MUG to understand site usability and use in both web and wireless contexts

Acknowledgments

The authors thank the senior editor, Allen Lee, the associate editor, and the three reviewers for their many excellent comments and suggestions that greatly improved this manuscript. The authors also thank Tracy Ann Sykes for her copy editing of earlier versions of this paper and Hillol Bala for proofreading. We also thank Jennifer Syverson for patiently answering our inquiries about the paper. Last but not least, the authors thank Jan DeGross for her tireless efforts typesetting our paper.

References

Agarwal, R., and Venkatesh, V. "Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability," *Information Systems Research* (13:2), 2002, pp. 168-186.

Ballard, B., and Miller, B. "WML Style Guide for the Phone. com 4. X Browser™," Sprint PCS White Paper, March 2001 (available online at http://developer.openwave.com/docs/carrier_sg/WMLStyleGuide1.pdf, accessed December 26, 2003).

Barnes, S., and Vidgen, R. "An Integrative Approach to the Assessment of E-commerce Quality," *Journal of Electronic Commerce Research* (3:3), 2002, pp. 114-127.

Barnes, S., Liu, K., and Vidgen, R. "Evaluating WAP News Sites: The WebQual/m Approach," in *Proceedings of the 9th European Conference on Information Systems*, S. Smithson, J. Gricar, M. Podlogar, and S. Avgerinou (eds.), Bled, Slovenia, 2001, pp. 344-355.

Battig, W. F. "Parsimony in Psychology," *Psychological Reports* (11:2), 1962, pp. 555-572.

Benbasat, I., and Zmud, R. W. "The Identity Crisis Within the IS Discipline: Defining and Communicating the Discipline's Core Properties," *MIS Quarterly* (27:2), 2003, pp. 183-194.

Buchanan, G., Farrant, S., Jones, M., Thimbelby, H., Marsden, G., and Pazzani, M. "Improving Mobile Internet Usability," in *Proceedings of the 10th ACM WWW Conference*, ACM Press, New York, 2001, pp. 673-680.

Castelhano, M. S., and Muter, P. "Optimizing the Reading of Electronic Text Using Rapid Serial Visual Presentation," *Behaviour & Information Technology* (20:4), 2001, pp. 237-247.

Chaiken, S., Duckworth, K. L., and Darke, P. T. "When Parsimony Fails…," *Psychological Inquiry* (10:2), 1999, pp. 118-123.

Chin, W. W., Marcolin, B. L., and Newsted, P. R. "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study," *Information Systems Research* (14:2), 2003, pp. 189-217.

Cook, T. D., and Campbell, D. T. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Houghton Mifflin Company, Boston, 1979.

Czerwinski, M., Larson, K., and Robbins, D. "Designing for Navigating Personal Web Information: Retrieval Cues," in *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting,* October 5-9, 1998, pp. 458-462

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science* (35:8), 1989, pp. 982-1003.

DeLone W. H., and McLean E. R. "Information System Success: The Quest for the Dependent Variable," *Information Systems Research* (3:1), 1992, pp. 60-95.

Duchnicky, R. L., and Kolers, P. A. "Readability of Text Scrolled on Visual Display Terminals as a Function of Window Size," *Human Factors* (25), 1983, pp. 683-892.

Ericsson Inc. "How to Make a Web Site Mobile: Extending Web Services with a Successful Mobile Service," White Paper, December 2002 (available online at http://www.ericsson.com/mobilityworld/developerszonedown/downloads/docs/wap/how_to_make_a_web_site_mobile.pdf; accessed December 26, 2003).

Ghorpade, J., Hattrup, K., and Lackritz, J. R. "The Use of Personality Measures in Cross-Cultural Research: A Test of Three Personality Scales Across Two Countries," *Journal of Applied Psychology* (84:5), 1999, pp. 670-679.

Han, S. H., and Kwahk, J. "Design of a Menu for Small Displays Presenting a Single Item at a Time," in *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, Nashville, October 24-28, 1994, pp. 360-364.

Huizingh, E. K. R. E. "The Content and Design of Web Sites: An Empirical Study," *Information and Management* (37:3), 2000, pp. 123-134.

Jackson, P., Favier, J., and Stagia, I. *Segmenting Europe's Mobile Consumers*, Forrester Research, Cambridge, MA, 2002.

Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., and Buchanan, G. "Improving Web Interaction in Small Screen Displays," in *Proceedings of the 8th World Wide Web Conference*, Toronto, May 11-14, 1999, pp. 51-59.

Karat, J. "Evolving the Scope of User-Centered Design," *Communications of the ACM* (40:7), 1997, pp. 33-38.

Kaikkonen, A., and Ruto, V. "Navigating in a Mobile XHTML Application," *SIGCHI Letters* (5:1), 2003, pp. 329-336.

Karkkainen, L., and Laarni, J. "Designing for Small Display Screens," in *Proceedings of the 2nd Nordic Conference on Human-Computer Interaction*, October 19-23, 2002, pp. 227-230.

Keeker, K. "Improving Web-Site Usability and Appeal: Guidelines Compiled by MSN Usability Research," Microsoft.com, July 24, 1997 (available online at http://msdn.microsoft.com/archive/default.asp?url=/archive/en-us/dnarplan/html/IMPROVINGSITEUSA.asp; accessed August 12, 2002).

Klein, L. R. "Evaluating the Potential of Interactive Media Through a New Lens: Search Versus Experience Goods," *Journal of Business Research* (41:3), 1998, pp. 195-203.

Koufaris, M. "Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior," *Information Systems Research* (13:2), 2002, pp. 205-223.

Kraut, R., Mukhopadhyay, T., Szczypula, J., Kiesler, S., and Scherlis, W. "Communication and Information: Alternative Uses of the Internet in Households," *Information Systems Research* (10:4), 1999, pp. 287-303.

Lederer, A., Maupin, D. J., Sena, M. P., and Zhuang, Y. "The Role of Ease of Use, Usefulness and Attitude in the Prediction of World Wide Web Usage," in *Proceedings of the 1998 ACM SIGCPR Conference on Computer Personnel Research*, R. Agarwal (ed.), ACM Press, New York, 1998, pp. 195-204.

Lee, A., and Baskerville, R. L. "Generalizing Generalizability in Information Systems Research," *Information Systems Research* (14:3), 2003, pp. 221-243.

Lohse, G. L., and Spiller, P. "Electronic Shopping," *Communications of the ACM* (41:7), 1998, pp. 81-88.

Lohse, G. L., and Spiller, P. "Internet Retail Store Design: How the User Interface Influences Traffic and Sales," *Journal of Computer-Mediated Communication* (5:2), 1999 (available online at http://jcmc.indiana.edu/vol5/issue2/lohse.htm).

Nielsen, J. "User Interface Directions for the Web," *Communications of the ACM* (42:1), 1999, pp. 65-72.

Nielsen, J. *Designing Web Usability*, New Riders, Indianapolis, IN, 2000.

Osborn, C. "Customer Retention: Can Wireless Data Make 'em Stay Put?," in *The Future of Wireless: Business Strategies, Broadband Technologies, and Network Operations*, International Engineering Consortium (IEC), Chicago, IL, 2003, pp. 311-320.

Palmer, J. W. "Website Usability, Design, and Performance Metrics," *Information Systems Research* (13:2), 2002, pp. 151-185.

Plouffe, C. R., Hulland, J. S., and Vandenbosch, M. "Research Report: Richness Versus Parsimony in Modeling Technology Adoption Decisions—Understanding Merchant Adoption of a Smart Card-Based Payment System," *Information Systems Research* (12:2), 2001, pp. 208-222.

Ramsay, M., and Nielsen, J. *WAP Usability: Déjà vu: 1994 All Over Again*, Nielsen Norman Group, Fremont, CA, December 2000 (available online at http://www.NNgroup.com/reports/wap).

Ranganathan, C., and Ganapathy, S. "Key Dimensions of Business-to-Consumer Web Sites," *Information and Management* (29:6), 2002, pp. 457-465.

Ressler, R. H. "Some Justifications for Parsimony in Psychology," *Psychological Reports* (13:3), 1963, pp. 643-646.

Sadeh, N. *M-Commerce: Technologies, Services, and Business Models*, John Wiley & Sons, Inc., New York, 2002.

Seddon, P. B. "A Re-Specification and Extension of the DeLone and McLean Model of IS Success," *Information Systems Research* (8:3), 1997, pp. 240-253.

Serco. *Designing WAP Services: Usability Guidelines*, Serco Usability Services, London, 2000 (available online at http://www.usability.serco.com/research/suswapguide.pdf; accessed March 12, 2004).

Settles, C. *Wireless, Inc.*, American Management Association, New York, 2002.

Sivo, S. A., and Willson, V. L. "Is Parsimony Always Desirable? Identifying the Correct Model for a Longitudinal Panel Data Set," *Journal of Experimental Education* (66:3), 1998, 249-255.

Straub, D. "Validating Instruments in MIS Research," *MIS Quarterly* (13:2), 1989, pp 147-169.

Straub, D., Limayem, M., and Karahanna, E. "Measuring System Usage: Implications for IS Theory Testing," *Management Science* (41:8), 1995, pp. 1328-1342.

Taylor, S., and Todd, P. A. "Understanding Information Technology Usage: A Test of Competing Models," *Information Systems Research* (6:4), 1995, pp. 144-176.

van de Vijver, F., and Leung, K. *Methods and Data Analysis for Cross-cultural Research*, Sage Publications, Thousand Oaks, CA, 1997.

Venkatesh, A., and Vitalari, N. P. "An Emerging Distributed Work Arrangement: An Investigation of Computer-Based Supplemental Work at Home," *Management Science* (38:12), 1992, pp. 1687-1706.

Venkatesh, V. "Creation of Favorable User Perceptions: The Role of Intrinsic Motivation," *MIS Quarterly* (23:2), 1999, pp. 239-260.

Venkatesh, V. "Determinants of Perceived Ease of Use: Integrating Perceived Behavioral Control, Computer Anxiety and Enjoyment into the Technology Acceptance Model," *Information Systems Research* (11:4), 2000, pp. 342-365.

Venkatesh, V., and Brown, S. A. "A Longitudinal Investigation of Personal Computer Adoption in Homes: Adoption Determinants and Emerging Challenges," *MIS Quarterly* (25:1), 2001, pp. 71-102.

Venkatesh, V., and Davis, F. D. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science* (45:2), 2000, pp. 186-204.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. "User Acceptance of Information Technology: Toward A Unified View," *MIS Quarterly* (27:3), 2003, pp. 425-478.

Venkatraman, N., and Grant, J. "Construct Measurement in Organizational Strategy Research: A Critique and Proposal," *Academy of Management Review* (11:1), 1986, pp. 71-87.

Yankee Group. "Mobile User Survey Results Part 1: Will Next Generation Data Services Close the Value Gap," Yankee Group Research Inc., Boston, 2002.

About the Authors

Viswanath Venkatesh is professor and the first holder of the George and Boyce Billingsley Chair in Information Systems at the Walton College of Business, University of Arkansas. Prior to joining Arkansas, he was on the faculty at the University of Maryland. He received his Ph.D. from the University of Minnesota. His research focuses on understanding technology in organizations and homes by focusing on electronic commerce, end-user training, user acceptance of new technologies, and gender and age differences in technology adoption and use. His research has been published in leading information systems, organizational behavior, and psychology journals. *MIS Quarterly* named him "Reviewer of the Year" in 1999. He has served on or is currently serving as an associate editor on the board of *MIS Quarterly, Information Systems Research*, *Journal of the AIS,* and *Decision Sciences*.

V. Ramesh is an associate professor in the Information Systems Department and Ford Motor Company Teaching Fellow at Indiana University's Kelley School of Business. He is the director of the MS in Information Systems program. He has published over 30 papers in leading journals, including *MIS Quarterly*, *ACM Transactions on Information Systems*, *Communications of the ACM*, *Journal of Management Information Systems*, *Information Systems*, and *IEEE Expert*, as well as edited books and conference proceedings. He is a senior editor for *e-Service Journal* and serves on the editorial board of *Journal of Database Management*. His research interests are in data modeling, heterogeneous databases, virtual teams and groupware, usability in mobile systems, and software engineering. His teaching interests include database design, object-oriented design and programming (Java and VB.NET), Web applications development (JSP/Servlets, ASP.NET), IT audit and e-business infrastructure.