Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Web Site Usability, Design, and Performance Metrics

Jonathan W. Palmer,

To cite this article:
Jonathan W. Palmer, (2002) Web Site Usability, Design, and Performance Metrics. Information Systems Research 13(2):151-167. http://dx.doi.org/10.1287/isre.13.2.151.88

Web Site Usability, Design, and Performance Metrics

Jonathan W. Palmer
University of Maryland, R. H. Smith School of Business, Decision and Information Technologies, 4348 Van Munching Hall, College Park, Maryland 20742-1871
jpalmer@rhsmith.umd.edu

Web sites provide the key interface for consumer use of the Internet. This research reports on a series of three studies that develop and validate Web site usability, design and performance metrics, including download delay, navigability, site content, interactivity, and responsiveness. The performance metric that was developed includes the subconstructs user satisfaction, the likelihood of return, and the frequency of use.

Data was collected in 1997, 1999, and 2000 from corporate Web sites via three methods, namely, a jury, third-party ratings, and a software agent. Significant associations between Web site design elements and Web site performance indicate that the constructs demonstrate good nomological validity. Together, the three studies provide a set of measures with acceptable validity and reliability. The findings also suggest lack of significant common methods biases across the jury-collected data, third-party data, and agent-collected data.

Results suggest that Web site success is a first-order construct. Moreover, Web site success is significantly associated with Web site download delay (speed of access and display rate within the Web site), navigation (organization, arrangement, layout, and sequencing), content (amount and variety of product information), interactivity (customization and interactivity), and responsiveness (feedback options and FAQs).

(*e-Commerce, Web Metrics, or Measurement*; *Web Site Usability*; *Design and Performance Constructs*; *Construct Validity*; *Nomological Validity*)

Introduction

Over the next few years, the World Wide Web (the Web) is expected to increase by a factor of 20, growing to 200 million sites by 2005. The number of actual Web pages will increase even more, with existing Web sites continuing to add pages. Indeed, Nielsen (2000) projects a growth to 50 billion pages by 2005. Given the size of the phenomenon, the measure of what users want in a Web site is an important area of study because the Web site is a primary user interface for net-enabled business (Straub and Watson 2001), information provision, and promotional activities (Alba et al. 1997, Jarvenpaa and Todd 1997, Schubert and Selz 1998). Developing sites that are responsive to user needs is critical for all site designers and managers (Price 1997). For Web page owners to be successful and for users to be satisfied, Web sites need to consider usability and other design criteria (Nielsen 2000, Pearrow 2000, Shneiderman 1998). To facilitate such design work, therefore, psychometrically and methodologically acceptable metrics need to be established.

User interaction with sites that exhibit higher than normal usability have been significantly associated with performance improvement (Took 1990). Nielsen (2000) reports a nine times performance improvement of user success based on user-centered navigability. Poor interface design has been a key element in a number of high profile site failures (Buschke 1997, Chain Store Age 1997).

1047-7047/02/1302/0151$05.00
1526-5536 electronic ISSN

Web site design can benefit from the application of usability principles (Nielsen 2000, Shneidermann 1998), media richness theory (Palmer and Griffith 1998, Schubert and Selz 1998, Trevino et al. 1990), and marketing concepts (Hoffman and Novak 1996). The question remains as to how to develop a reliable and valid set of metrics for Web site design and performance. The current research reports on a set of three studies of Web sites aimed at developing such a set of design and performance metrics.

Need for Metrics

Metrics help organizations generate more effective Web sites and provide measures that managers understand and that academics can replicate and analyze. To provide practical value, metrics should identify frequency of measurement, frequency of review, source of data, rationale for introducing the measure, who will act on the data, and the purpose of the measure (Neely 1998). For scientific, quantitative rigor, metrics should exhibit, at a minimum, construct validity and reliability (Straub 1989, Cook and Campbell 1979).

Measurement is the process by which numbers or symbols are assigned to attributes of an entity in a way that helps describe the entity according to clearly defined rules (Fenton 1994, Nunnally 1978, Churchill 1979). There are typically internal attributes of a product (such as a Web site) that can be measured purely in terms of the product; there are also external attributes, which, in turn, are measured in relation to how the product interacts with other entities in the environment. These are related. As Fenton (1994) says: "[W]e need to measure internal attributes to support the measurement of external attributes" (p. 204). The use of multi-item measures allows for better measurement, tending to increase reliability and decrease measurement error (Churchill 1979, Cook and Campbell 1979).

For effective measurement, the measurement activity should have clear objectives and identify types of attributes that can be measured and appropriate scales. Empirical relations for an attribute should be identified in advance and the measurement scale must be meaningful. A continual attempt to refine and improve scales based on data analysis leads to clarification and improved evaluation of the attribute (Fenton 1994).

The purpose of this research is to identify appropriate metrics for usability, design, and performance constructs to be used in studying consumer-focused Web sites. Web sites differ dramatically in their level of sophistication, ranging from heavily text-based sites to those replete with multimedia (Ho 1997, Griffith and Palmer 1999, Rose et al. 1999). Each site is subject to scrutiny through fundamental questions such as: Which design elements of a Web site are most critical? What elements of the design impact user response to the Web site? What are appropriate measures for Web site success? These are some of the research questions that drive this set of studies.

Prior Research

Usability and Design

The usability and design of Web sites has received attention in the human computer interaction (HCI) literature as well as in Web-specific usability research. Usability has typically taken an engineering approach in an attempt to identify a set of principles and common practices that will insure usability is an outcome of system design (Nielsen 1993, Pearrow 2000, Shneidermann 1998).

Prior to widespread use of the Web, usability of information systems was equivalent to a set of design principles, articulating five key elements including: (1) consistency of the interface, (2) response time,[1] (3) mapping and metaphors, (4) interaction styles, and (5) multimedia and audiovisual (Nielsen 1993). Consistency suggests the need for common placement of navigational tools, such as buttons and bars. Response time focuses on the speed with which the system provided a response to user activity. Mapping and metaphors in usability stresses navigation from place to place within the system and the adoption of specific metaphors (Przystupa 1993), such as shopping carts, to aid in user activity within the system. Interaction styles concentrates on system messages that are generated in response to user activity. The fifth usability element is the degree to which multimedia capabilities are incorporated into the system design.

[1]Consistent with Rose et al. (1999) the term "download delay" has been adopted for this construct throughout.

As the Web became an increasingly essential interface, usability research began to focus more specifically on extending the basic usability principles into the Web environment (Nielsen 2000, Shneidermann 1998). Nielsen (2000) extended these design principles for Web design to include: (1) navigation, (2) response time, (3) credibility, and (4) content. This suggests easy-to-use navigation, frequent updating, minimal download times, relevance to users, and high-quality content that also takes advantage of capabilities unique to the online medium (Nielsen 1993).

Navigation is an important design element, allowing users to acquire more of the information they are seeking and making the information easier to find (Machlis 1998b). Thus, a key challenge in building a usable Web site is to create good links and navigation mechanisms (Radosevich 1997). Graphical design, layout, and actual content are prime components in making the page easier to use (Rasmussen 1996). Text links are vital; navigation and content are inseparable; and key areas are navigational structure, searching, readability, and graphics (Spool 1997).

As Web sites have matured, some homogenization has occurred, with de facto standards beginning to emerge (for example, the use of the basic catalog of product presentations and navigation protocols). The nature of design has also been influenced by changes in technology, but because these changes are usually not disruptive, basic design principles tend to endure (Pearrow 2000).

Research has been ongoing in identifying approaches to improve the ease of use of the Web (Levi and Conrad 1996, Boling 1995). Results often focus on the download delay, success in finding a page, and organization of the information gathered during a Web session (Pitkow and Kehoe 1996). A Web site high in usability should generate a desirable perception of its use and an intention to use the site. Usability includes consistency and the ease of getting the Web site to do what the user intends it to do, clarity of interaction, ease of reading, arrangement of information, speed and layout. Appropriate design of user interfaces includes organization, presentation, and interactivity (Shneiderman 1998).

Selz and Schubert (1997) and Schubert and Selz (1998) develop a Web assessment model that includes the identification of "media inherent characteristics." These characteristics include hypermedia presentation (content structure, user interface and potential for combinations), time (availability and contact possibilities), ubiquity (availability and system performance), expert systems (personalization, possible product combination, configuration), and interactivity (customer profile, user determining look of interface).

Prior research also provides insights into activities supported by Web sites. A key capability of the Internet is its capacity to support greater interactivity for users. Consumers using the Internet to gather information, to secure product information, and to purchase goods and services are influenced by the interactivity of the Web site (Jarvenpaa and Todd 1997, Alba et al. 1997). Another set of characteristics that users consider when responding to Web sites is its media richness.

Media Richness

Media richness refers to a medium's relative ability to convey messages. Media have a range of "richness" (i.e., "the ability of information to change understanding within a time interval," Daft and Lengel 1986, p. 560) from "very rich" to "very lean." On the basis of these differences, channels can be arrayed along a continuum describing their relative richness, which has been labeled the "media richness continuum" (Daft et al. 1987, Rice 1992, Trevino et al. 1987).

The Web provides users with a number of levels of richness. As a medium, it can range from text-based to multimedia. This array of functionalities makes the Web site a medium capable of adopting elements of other media such as memos, letters, fax, audio and videoconferencing, and e-mail. Thus, the range of channel capacities make a Web site difficult to position as a medium along the continuum initially proposed by Daft and Lengel (1986). These original classifications, which included face to face, telephone, personal documents, impersonal documents, and numerical documents, have been extended to include e-mail (Markus 1994, Trevino et al. 1990, Huang et al. 1998, Rice and Shook 1990, Rice et al. 1989, Schmitz and Fulk 1991), fax (Donabedian et al. 1998), voice mail (D'Ambra and Rice 1994, El-Shinnawy and Markus 1998, Straub and Karahanna 1997), audio/video (Suh 1999), and "new

media" computer-mediated and video communication (Dennis and Kinney 1998). According to media richness theory, the multimedia interactive format should provide capabilities richer than the text and photographs of sales brochures and catalogs (Palmer 1997). Web sites can take advantage of this multimedia capability to enhance the medium richness and can utilize "frames" to provide access to multiple pages simultaneously (Palmer and Griffith 1998).

The specific influence of a given medium is often dependent on the task being performed (Daft and Lengel 1986; Rice 1992, 1993). Some suggest a difference in media perceptions across cultures (Rice and D'Ambra 1998, Ross 2001) and the influence of use experience in perceptions of media richness (Carlson and Zmud 1999). Media richness also suggests a Web site can be differentiated on the richness of product information and responsiveness to the user. Providing high-quality information directly related to a product's salient attributes enhances consumer response (Alba et al. 1997). Finding information that is of high quality within the computer-mediated context is also an important element (Hoffman et al. 1995, Dickson 2000).

Considering the background offered by these literatures, it is clear that usability and design research suggests several interface elements for Web site development. Moreover, media richness adds specific media-inherent capabilities and information content that need to be included in an initial set of potential Web metrics.

Web Site Metrics and Hypothesis Development

The application of usability principles and richer media in Web site design should result in more successful Web sites. Metrics that capture these key elements of usability, design, and media richness should aid Web site designers. Metrics include measures to capture both design characteristics and Web site outcomes and, based on prior research modeling, lead to a set of hypotheses.

Outcome Metrics: The Dependent Variable

As with any information system, there are likely to be multiple dimensions of Web site success. Frequency of use, user satisfaction, information quality, system quality, and impact are all elements of success (DeLone and McLean 1992). User participation, involvement, and attitude are also important dimensions and the measurement of these items has been well developed in the IS literature (Barki and Hardwick 1994). These measures are interdependent. For the purposes of this research, our metrics for Web site success concentrated heavily on frequency of use, user satisfaction, and intent to return as key measures of the user's success with the Web site.

Commonly applied usage measures, such as number of times a Web page is viewed or visited, have generated conflicting results. Page views or visits fail to provide adequate insight into the ultimate success of a Web site (Picarille 1997), but Web site traffic measures have also been found to be a good surrogate for Web site efficiency as a measure of success (Alpar et al. 2001). Most Web usage tracking tools measure Web hits and record a visitor's domain name of origin and the most frequently called pages (Bort 1997). In many cases these measures are aimed at activity on the Web site in terms of visitor traffic, page views, and return visits and are often tracked through the use of cookies (Cohen 1997, Machlis 1998a, Stone 1999). These traffic measures have been used because they are easy to capture, but they are very often deemed to be inadequate (Rose et al. 1999).

Beyond merely tracking usage, Griffith et al. (2001) integrate the marketing (McQuarrie and Munson 1991, Munch et al. 1993, Petty and Cacioppo 1979) and IS literatures (Barki and Hardwick 1994) by proposing the construct *user satisfaction* for the Web site interface. Ease of use and usefulness are also important candidates for the measurement of success (Adams et al. 1992).

Armstrong and Hagel (1996) go on to suggest the need for developing a sense of community and loyalty. While there are difficulties in measurement of loyalty per se, the measurement of frequency of use and likelihood of return may capture some of these elements (Niccolai 1997). This suggests two additional candidates for measuring Web site performance, namely, *frequency of use* and *likelihood of return*. Web site outcomes can be influenced by a number of usability and

design elements that will serve as the independent variables in this research.

Usability and Design Metrics: Independent Variables

There have been some actions at an industry level to identify a Web site performance index, which measures technical dimensions such as availability and download delay (speed of access). Bizrate ⟨web.bizrate .com⟩ captures information on the operational efficiency of sites, compiling ratings from individual consumers to develop its metrics for convenience and speed of ordering. Metrics based on transaction processing time and rate, service failures, download delay or "user response time" (Barney 2000, Messmer 1999, Wilson 1999), and site reliability (Berry, 1999) have also been suggested. Finally, Evans and Wurster (2000) and Rose and Straub (2001) suggest that operational efficiencies on Web sites should include *download delay* or speed of response.

Usability and design research suggests *navigability and organization* are potential key variables (Nielsen 2000, Shneidermann 1998). Media richness suggests *content richness* and *responsiveness*. Both design and media richness suggest a concept of user *interactivity* with the system or medium (Palmer and Griffith 1998, Trevino et al. 1990). All of these are strong potential candidates as independent variables.

Research Model and Hypothesis Development for Nomological Validity

Elements of Web site design have been related to success outcomes in prior literature and the elaboration of this research model will enable us to test for the nomological validity of design metrics later in the paper. Hypotheses expressing each of these relationships follow.

First, the focus of this research is on elements of a Web site that can be controlled. For example, telecommunications infrastructure that might influence a user's speed of access is beyond the designer's control, and will not be considered here. However, Web site designers can choose not to include slow loading elements such as longer audio or video clips, reducing initial access time (Levine 1996), and this element we are modeling. Length of wait is important, as users are often unwilling to wait more than a handful of seconds for a response (Shneiderman 1998). For Web sites, response time is actually the *download delay*[2] for each of these activities. That is, it is the initial request for access to the page and then each subsequent request for changing pages within the site (Rose et al. 1999). This suggests a basic download delay hypothesis articulated in the literature (Rose and Straub 2001).

Hypothesis 1. *Web sites exhibiting lower download delay will be associated with greater perceived success by site users.*

Usability researchers suggest that organization and navigation is important to outcomes (Nielsen 2000). Web site elements include design, layout, sequencing, and arrangement (Schonberg et al. 2000). *Navigability* is defined as the sequencing of pages, well organized layout, and consistency of navigation protocols. This suggests our second hypothesis.

Hypothesis 2. *More navigable Web sites will be associated with greater perceived success by site users.*

User manipulation and utilization of the information provided via Web sites is strongly influenced by interface design. The ability to provide a personalized, customized interaction for the user allows Web site design that differentiates product and service offerings (Palmer and Griffith 1998). *Interactivity* includes the ability to customize the site's look, feel, and content as well as provide interaction with the user. This suggests a third hypothesis.

Hypothesis 3. *Higher interactivity in Web sites will be associated with greater perceived success by site users.*

Some of the most valued metrics revolve around customer reaction to the site (Berry 1999). User interface measures suggested by Shapiro and Varian (1999) include feedback features and functions. Evans and Wurster (2000) suggest feedback and access to previously asked questions as important. Responsiveness is a key consumer issue when shopping on the Web (Jarvenpaa and Todd 1997). *Responsiveness* is defined as

[2]Usability literature has typically used the term "response time" (Shneiderman 1998, Nielsen 2000). As noted earlier, this research adopts "download delay" (Rose et al. 1999) as more accurate terminology to differentiate this construct from the responsiveness construct.

the presence of feedback to users and the availability of response from the site managers. This prior research suggests the following hypothesis.

HYPOTHESIS 4. *More responsive Web sites will be associated with greater perceived success by site users.*

Media richness theory (Daft and Lengel 1986) suggests that the quality, accuracy, and reliability of the information exchanged across a medium are critical. Key Web site capabilities include comprehensiveness and completeness of information (Shapiro and Varian 1999). Evans and Wurster (2000) highlight richness and quality of information. Content quality and variety are also key consumer measures when shopping on the Web (Jarvenpaa and Todd 1997).

Reflecting this interest in content, the Gomez Performance Tool ⟨web.gomez.com⟩ includes sundry metrics for onsite resources: content, tools, and products. Bizrate (web.bizrate.com) also captures information on breadth/depth of products offered, product information quantity, and quality and relevance. Therefore, our construct *content* includes the amount and variety of content as well as the use of text, graphics, and multimedia. This suggests the content hypothesis:

HYPOTHESIS 5. *Higher quality content in Web sites will be associated with greater perceived success by site users.*

Proposed metrics for each of the five independent variables used in the research model as well as dependent variables are presented in Table 1. The three methods used to collect measures—jury, agent, and Alexa—are also indicated in the table.

Methodology

To collect usability and design data and examine nomological validity through testing the theoretical relationships, we analyzed actual Web sites of large, publicly held firms. Three studies were completed during the three-year period from early 1997 to early 2000. In that other researchers looking at Web usability surveyed Web users (Lederer et al. 1998, Griffith et al. 2001), one of our methods in Study 1 was a jury of respondents assuming the role of Web site users. Each succeeding study built on metrics identified and refined in the prior study and exercised these metrics through the use of one additional data gathering method. Thus, the second study incorporated third-party data besides jury data, and the third study added agent-gathered data to the third-party and jury data. An overview and comparison of the three studies and methods is provided in Table 2.

The challenge was to identify and measure the best characteristics and design of Web interfaces to insure effective performance and user satisfaction in such a way so that other researchers could compare, replicate, and utilize the model. Initial construct validity proved to be difficult, because the literature was not fully developed in 1997 and Web technology was continually evolving. Consequently, the instrument was enhanced in 1999 via new data collection methods.

Instrumentation

Straub (1989) suggests validating instruments through pretesting and/or piloting using previously validated instruments wherever possible, whereas technical validation can be undertaken to ground central constructs. Boudreau et al. (2001) argue that use of existing instruments is primarily for purposes of efficiency, but is still well regarded as a methodological approach. Indeed, utilization of the existing literature as a basis for conceptualizing and specifying constructs supports specification of a set of potentially strong metrics (Churchill 1979). Thus, prespecified constructs were utilized for the instrument development where possible. Involvement with the Web site was operationalized using McQuarrie and Munson (1991). User evaluation of a Web site was adapted from the scale developed originally by Petty and Cacioppo (1979).

When no scales existed, the instrument was developed from items in the literatures of media richness (Daft and Lengel 1986), usability (Nielsen 1993), design (Schubert and Selz 1998, Selz and Schubert 1997), and from early work in electronic commerce (Palmer 1997, Hoffman and Novak 1996). The initial iteration of the instrument in 1997 included pretesting and a pilot instrument in a classroom experiment. Development of the instrument followed existing theory in the absence of established measures (Boudreau et al. 2001). In an attempt to generate unidimensional responses (Bagozzi 1981), semantic differential scales were used

Table 1 Constructs, Items, and Sources

Construct	Measures	Method	Sources
Hypothesis 1: Download Delay	Initial Access Speed, Speed of Display Between Pages; Alexa Speed	Jury: Seven-Point Scale Alexa: Five-Point Scale WebL: Actual Time	Barney 2000 Messmer 1999 Pitkow and Kehoe 1996 Nielsen 1993 Rose et al. 1999
Hypothesis 2: Navigation/ Organization	Arrangement, Sequence, Links, Layout, Alexa Organization	Jury: Seven-Point Scale Alexa: Five-Point Scale	Boling 1995 Nielsen 1993 Randall 1997 Schubert and Selz 1998 Schonberg et al. 2000 Shneiderman 1998
Hypothesis 3: Interactivity	Customization Interactivity	Jury: Seven-Point Scale	Chain Store Age 1997 Alba et al. 1997 Jarvenpaa and Todd 1997 Shneiderman 1998
Hypothesis 4: Responsiveness	Feedback FAQ	Jury: Seven-Point Scale	Chain Store Age 1997 Shneiderman 1998
Hypothesis 5: Information/ Content	Amount of Information, Variety of Information, Word Count, Content Quality	Jury: Seven-Point Scale Alexa: Five-Point Scale WebL: Word Count	Berthon et al. 1996 Evans and Wurster 2000 Shapiro and Varian 1999 Shneiderman 1998
DV: Web Site Success	Satisfaction, Likelihood of Return, Frequency of Use,	Jury: Seven-Point Scale	DeLone and McLean 1992, Rose et al. 1999 Armstrong and Hagel 1996, Niccolai 1997

to capture individual items in several instances (including Web site success, ease of navigation, graphical content, and page layout).

The final iteration of the instrument (utilized in 1999 and 2000 and attached as an appendix) showed relatively small changes from a technology-specific measures (e.g., frames, push-to-talk) to direct measures of usability and design functionality. Such changes as were made were consistent with the advice of Churchill (1979) to eliminate items that do not share equally in the common core of a construct.

Data Collection and Analysis

The first study took place in early 1997. A group of 35 respondents made up of undergraduate and MBA students evaluated 250 randomly selected firms of the U.S. *Fortune* 500 corporate Web sites over a two-week period. The *Fortune* 500 represents firms that are most likely to have the capacity for supporting use of Web sites and are the most probable early adopters of leading technological capabilities. Even so, few sites at the time were actually using all available aspects of extant Web technology. Jury data was factor analyzed to examine constructs and regressed on performance to test the hypotheses.

In early 1999, an updated Web site analysis was performed on 250 sites randomly selected from the *Global Fortune* 1000 with a jury of 30 respondents, including undergraduate and MBA students. As noted above, constructs used in the 1997 study were extended through a series of items to tap into the functionality rather than the technology of the Web site. This second study also utilized measures from the Alexa third-party source in evaluating the sites. Alexa retrieves contact information from official Internet registration organizations around the world, such as InterNIC and

Table 2 Overview of Studies: Characteristics and Metrics

	1997	1999	2000
		Characteristics	
Focus	Technology	Site Capabilities	Site Capabilities
Level of Adoption	Early Adopters	More Mature Uses	Greater Maturity
Web Sites	250 of *Fortune* 500	250 of *Global* 1000	250 of *Global* 1000
Nature of Constructs	Muddy Constructs	Better Specified Constructs	Even Better Specified Constructs
Method	Jury	Jury	Jury
		Third-Party Information	Third-Party Information
			Agent Data Collection
		Metrics	
Download Delay	Access Speed, Display Rate	Access Speed, Display Rate	Access Speed, Display Rate
Navigation	Navigability—Frames	Sequence, Layout, Arrangement	Sequence, Layout, Arrangement
Content	Text/Graphics	Content	Content
Interactivity	A/V Multimedia	Interactivity	Interactivity
Responsiveness	FAQ, Feedback	Responsiveness	Responsiveness
Site Success	Likely to Return, Frequency of Use, User Satisfaction	Likely to Return, Frequency of Use, User Satisfaction	Likely to Return, Frequency of Use, User Satisfaction

RIPE, as well as from individual Web site and home page owners. Alexa's speed rating indicates a site's effective transfer rate measured in kilobits/second during Alexa's periodic crawls of the Internet. Alexa's quality and organizational ratings were derived from user reviews of Web site content.

Study 2 captured more generalizable measures of navigability, information content, download delay, responsiveness, and interactivity. This study also confirmed a set of dependent variable measures for user satisfaction, likelihood of return, and frequency of use. The resulting data set included jurors' completed analysis forms and specific Alexa ratings for each site. Jury and Alexa data were factor analyzed to examine constructs, and regression analysis to examine nomological validity, which was, in effect, a testing of the hypotheses.

The third study was completed in January 2000. We evaluated 250 *Global Fortune* 1000 Web sites randomly selected from the 750 not utilized in the 1999 study. An instrument identical to that used in Study 2 was used to collect data from a jury of 35 respondents including MBA and executive MBA students. To more closely control the data collection and to avoid differences in connect time to Web sites, the evaluation process was conducted in a university computer lab over a five-day period. The study also collected the same Alexa data. Additionally, this third study included a third data collection method, a software agent programmed in WebL. This provided a third method for examining the targeted Web sites.

Web Language ⟨http://www.research.compaq.com/SRC/WebL/⟩ is a scripting language for automating tasks on the World Wide Web. It is an imperative, interpreted language that has built-in support for common Web protocols like HTTP and FTP and popular data types like HTML and XML. Figure 1 shows the agent gathering word counts for the IBM page.

This agent ran throughout a one-month period to assess download delays and word count during the five-day jury and Alexa data gathering period. It needs to be noted that this form of measuring download delay captures only part of the total time, specifically the server-side delay and the transmission delay (Rose et al. 1999). It does not capture the client-side delay, which was held constant, in this case at a baud rate of 56.6K.

Figure 1 Web Language Module for Word Count

```
create QueryURLList
create LogFile                          // e.g., "IBM 1214.log"
for every URL in QueryURLList do
  start logging import
  get StartPage
  if StartPage Available
    var txt = Text(page);
    var words = Size(Str_Search(txt, "[0-9a-zA-Z']+"));
  else
    add "n/a" to DataFileStr
  endfor
create DataFile                         // e.g., "IBM 1214.txt
save DataFileStr to DataFile
save ProcessTime to LogFile
end;
```

The complete data set for Study 3 included the jurors' completed analysis forms, specific Alexa ratings for each site, and the agent-collected data. Again, the data set was factor analyzed to examine constructs from the jury data, and regression analysis to test nomological validity and the hypotheses.

Instrument Validity

Since the instrument in Study 1 was more in the vein of a pretest of instrument validity, we next analyze the validity of the refined instrument used in Studies 2 and 3. The analysis includes reliability, convergent and discriminant validity (construct validity), and nomological validity as well as external validity or generalizability.

Reliability. Reliability, the extent to which an individual juror could answer the same question the same way each time, was assessed using Cronbach's α (1951). As recommended for exploratory work by Nunnally (1967), reliabilities in Study 1 were above 0.60 for four of the constructs, but some lower αs suggested that changes needed to be made in the instrumentation. Studies 2 and 3, with the final metrics, exhibited reliabilities from 0.70 to 0.96, and were acceptable by Nunnally's (1978) later standards.

Construct Validity. Construct validity improved across the three studies. First, jury inter-rater reliability improved across the studies, suggesting that the constructs were meaningful. In each of the studies, inter-rater reliability was analyzed via Cohen's (1960) coefficient Kappa, a stringent measure of inter-rater agreement. Maximum values of Kappa, as a function of observed agreement between jurors, were determined (Umesh et al. 1989). For all items, the average Kappa was greater than 0.75, which exceeds the suggested minimum of 0.70 (Landis and Koch 1977).

Convergent and Discriminant Validity (Construct Validity). Convergent validity is high when the measures of a construct from one measurement method correlate highly with other measures using different methods to measure the same construct (Campbell and Fiske 1959). Study 1 used a single measurement method, Study 2 used two methods, and Study 3 employed three methods. In Study 3 the correlation among those factors with multiple measurement methods was quite high in many cases and universally significant (see Table 3). The highlighted numbers suggest acceptable convergent validity on download delay (including jury rating of download delay, Alexa rating of download time, and WebL agent measure of download delay), navigation (jury and Alexa ratings), and content factors (jury rating, Alexa rating, and WebL agent word count).[3] Information content showed particularly strong convergence, as reflected in the high inter-method correlations.

Discriminant validity refers to the extent to which measures of different constructs, using different methods, are distinct from each other (Campbell and Fiske 1959). Discriminant validity increased from Study 1 to Study 3, with lower correlations among the four constructs, as shown in Table 3 (Churchill 1979).

Confirmatory factor analysis was used next to assess the convergent and discriminant validity of the items in Study 3. Using confirmatory factor analysis identifies a unique solution and provides a more rigorous approach to scale validation (Smith et al. 1996, Bagozzi 1983). Using LISREL, the confirmatory factor model for the constructs of navigation, content, interactivity, responsiveness, and site success provided an acceptable

[3]This is not strictly speaking an MTMM analysis (Campbell and Fiske 1959) but a simpler comparison of correlations to provide evidence for the methodological concept of convergent validity. A rigorous MTMM analysis would reveal some violations in the matrix, but the methodological literature is not clear on what are acceptable levels of these violations in any case (Alwin 1974). So, for the sake of parsimony of interpretation, we opted for a simpler, more straight-forward analysis.

fit to the data (χ^2 = 147.16, df = 55, AGFI = 0.91, RMR = 0.044) (Gefen et al. 2000).

Further Validation of Metrics: Nomological Validity

Nomological validity is the validation of constructs as correct predictors of an established "nomological network" of relationships (Campbell 1960). Nomological validity involves examining the modelled relationships between the latent constructs and if these are related as predicted by theory and/or prior literature, then the constructs themselves are equally validated (Bagozzi 1980, Straub et al. 1995, Boudreau et al. 2001). Regression analyses show that the design and usability items in Study 2 and Study 3 are positively associated with site success. This supports the claim of nomological validity of the metrics. The findings from all three of the studies are summarized in Table 4.

To organize the results of the three studies, the paper proceeds chronologically through the 1997, 1999, and 2000 data sets. Each study presents usability and design measures and compares their predicted relationship with Web site success.

Study 1

Results from the initial, exploratory study created metrics heavily based on the prevailing technologies. Factor analysis suggests that of the five key constructs, the first four show acceptable reliabilities (Nunnally 1967). They are: download delay (α = 0.75) navigability (α = 0.76), interactivity (α = 0.78), content/information richness (α = 0.67), and multimedia (α = 0.47). The factors show good construct validity in that the expected items "converge" into the proper factors, and they do not cross-load, i.e., they discriminate, as predicted (see Table 5).

In addition to the independent variables, the dependent variables loaded on a single construct of Web site

Table 3 Study 3 Correlations of Multimethod Measures

	Jury Download	Alexa Download	WebL Download	Jury Navigate	Alexa Navigate	Jury Content	Alexa Content
Alexa Download	**0.220****						
WebL Download	**0.784*****	**0.288****					
Jury Navigate	−0.035	0.156**	0.115*				
Alexa Navigate	0.206**	0.071	0.195**	**0.686*****			
Jury Content	0.030	−0.102*	0.203**	−0.066	0.156*		
Alexa Content	0.135*	−0.143**	0.178**	−0.006	0.096	**0.674*****	
WebL Content	0.274**	0.036	0.295**	0.270**	0.396***	**0.499*****	**0.497*****

*** Significant at 0.001; ** Significant at 0.01; * Significant at 0.05 (two-tailed).

Table 4 Summary of Findings

Hypothesis	Study 1	Study 2	Study 3
(1) Web sites exhibiting lower download delay will be associated with greater perceived success by site users.	Supported at 0.10 Level	Supported	Supported for Jury and Agent
(2) More navigable Web sites will be associated with greater perceived success by site users.	Supported	Supported	Supported
(3) Higher interactivity in Web sites will be associated with greater perceived success by site users.	Not Supported	Supported	Supported
(4) More responsive Web sites will be associated with greater perceived success by site users.	Not Measured	Supported	Supported
(5) Higher quality content in Web sites will be associated with greater perceived success by site users.	Supported	Supported	Supported

Table 5 Study 1—Factor Analysis for Jury Independent Variables

	Download Delay	Naviga-bility	Inter-activity	Content/Info Richness	Multi-media
Eigen Value	1.49	2.77	1.72	1.66	1.41
Variance Explained	10.21	19.81	12.27	11.87	10.05
Standardized Alpha	0.75	0.76	0.78	0.67	0.47
Access Time	0.68				
Display Rate	0.66				
Text		0.64			
Graphics		0.76			
Internal Links		0.76			
Navigability		0.66			
Colorful		0.74			
S/ware Download			0.95		
Free Software			0.95		
Free Evaluation			0.58		
FAQs			0.56		
Feedback			0.46		
Product Information				0.73	
Variety of Products				0.74	
Visually Attractive				0.73	
Audio					0.80
Video					0.74

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Factors with loadings < 0.40 are omitted.

success made up of likelihood of return (0.78), frequency of use (0.83), and satisfaction (0.91), and exhibited good reliability ($\alpha = 0.83$).

These same five factors were utilized in the regression analysis[4] presented in Table 6. Findings suggest that navigation and content are significantly associated with Web site success at the 0.05 α level. Neither interactivity nor multimedia was significantly associated with Web site success. These results may be influenced by the fact that in 1997, few sites had adopted technologies to support interactivity and multimedia. The first study findings do support Hypothesis 1 for download delay (but only at the 0.10 level), Hypothesis 2 (navigation), and Hypothesis 5 (content). Although this test did not yield across-the-board support for the nomological validity of the measures, it provided enough evidence to proceed with the next two studies.

[4]Factor scores for each of the five factors were obtained using the regression method and then used as variables in the regression model (Johnson and Wichern 1998).

Table 6 Study 1—Regression Analysis on Site Success

	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	*t*	Sig.
(Constant)	1.89	0.61		3.12	0.002
Download Delay	−0.48	0.27	0.37	−1.81	0.086
Navigation	0.21	0.08	0.27	2.75	0.007
Content	0.28	0.09	0.26	3.00	0.003
Interactivity	0.02	0.06	0.04	0.53	0.60
Multimedia	0.12	0.18	0.06	0.67	0.51

Adjusted R^2 is 0.44.

Reliable by Nunnally's 1967 standards, the download delay, navigation, content, and interactivity constructs explained sufficient variance (44%) to proceed with the development of new items and the elimination of others to improve the metrics.

Study 2

Instrumentation in the second study generated a more reliable set of metrics informed by the usability and media richness literatures. Reliabilities were: download delay ($\alpha = 0.96$), organization/navigation ($\alpha = 0.71$), information/content ($\alpha = 0.71$), interactivity ($\alpha = 0.70$), and responsiveness ($\alpha = 0.70$). The interactivity construct includes the Web site's use of customization and other interactive activities. The responsiveness construct in Study 2 incorporated the FAQ and feedback items from the interactivity construct in Study 1. As in Study 1, the dependent variables loaded on a single construct of Web site success (see Table 7) made up of likelihood of return (0.90), frequency of use (0.83), and satisfaction (0.90); performance was validated as a first-order construct with excellent reliability ($\alpha = 0.89$). These results suggest that different dimensions of Web site success reinforced each other and were closely associated in the minds of the jurors.

Building on the findings of the first study, the jury data was supplemented by information provided from the Alexa analysis tool. The five factors from the jury data and the three Alexa variables were then utilized in the regression analysis presented in Table 8. Findings suggest that all of the identified variables are significantly associated with Web site success. The second

study findings support all hypotheses, indicating that the constructs demonstrate nomological validity. They also suggest that the Alexa data and the jury data were very similar. Alexa data for speed shows a more significant result than the jury measure for download delay. Alexa data is captured over a longer period of time, so it may more closely capture the actual Web site download delay than the jury data.

Table 7 Study 2—Factor Analysis for Jury Independent Variables

	Download Delay	Organization/ Navigation	Information/ Content	Inter-activity	Respon-siveness
Eigen Value	1.45	3.85	1.84	1.41	1.37
Variance Expl	13.45	28.82	12.10	12.77	12.44
Stdzd Alpha	0.96	0.71	0.71	0.70	0.70
Access Time	0.89				
Display Rate	0.93				
Organization		0.76			
Sequence		0.75			
Arrangement		0.75			
Layout of Site		0.65			
Product Information			0.81		
Variety of Products			0.69		
Customization				0.75	
Interactivity				0.61	
FAQ					0.78
Feedback					0.59

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Factors with loadings < 0.40 are omitted.

Table 8 Study 2—Regression Analysis

	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
(Constant)	−0.71	0.20		3.51	0.000
Jury Download Delay	−0.007	0.04	−0.09	−1.75	0.08
Jury Organization/Navigation	0.21	0.02	0.21	9.41	0.000
Jury Content/Information	0.51	0.03	0.51	13.51	0.000
Jury Interactivity	0.27	0.08	0.27	3.36	0.001
Jury Responsiveness	0.29	0.02	0.29	12.71	0.000
Alexa Organization	0.27	0.03	0.27	7.14	0.000
Alexa Quality	0.009	0.02	0.09	4.39	0.000
Alexa Download	−0.15	−0.04	−0.28	−3.53	0.000

Adjusted R^2 is 0.71.

Study 2 resulted in dramatically improved metrics over Study 1. By introducing a second set of measures from Alexa, the constructs are validated symmetrically (Campbell and Fiske 1959). Study 3 sought to further validate the measures from Study 2.

Study 3

By identifying download delay, organization/navigation, information/content, interactivity, and responsiveness as the five key factors, Study 3 analysis yielded a factor solution consistent with our theoretical expectations and confirming the constructs from Study 2. Total variance explained by the five key factors was over 83%. In addition to clean factor loadings (demonstrating both convergent and discriminant validity), these four constructs also showed good reliability with standardized αs ranging from 0.72 to 0.96 (see Table 9).

The five jury factors, Alexa ratings, and agent-gathered data were entered into the regression presented in Table 10. WebL download delay was a two-item measure, incorporating initial access speed and display refresh speed, with a standardized α of 0.96. Findings suggest that all of the independent variables had a significant association with site success, with the

Table 9 Study 3—Factor Analysis for Jury Independent Variables

	Download Delay	Organization/ Navigation	Information/ Content	Inter-activity	Respon-siveness
Eigen Value	2.37	3.05	2.27	1.83	1.61
Variance Explained	18.22	22.72	17.64	14.69	12.48
Standardized Alpha	0.96	0.87	0.85	0.85	0.72
Access Time	0.95				
Display Rate	0.95				
Organization		0.85			
Sequence		0.87			
Arrangement		0.83			
Layout of Site		0.81			
Product Info			0.74		
Variety of Products			0.90		
Customization				0.85	
Interactivity				0.91	
FAQ					0.79
Feedback					0.89

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Factors with loadings < 0.40 are omitted.

Table 10 Study 3—Regression on Web Site Success

	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	*t*	Sig.
(Constant)	2.63	0.49		5.27	0.0001
Jury	0.008	0.01	−0.15	−6.88	0.0001
Download Delay					
Jury Organization/Navigation	0.12	0.05	0.128	2.59	0.01
Jury Information/Content	0.248	0.093	0.25	2.67	0.008
Jury Interactivity	0.004	0.021	0.04	2.01	0.05
Jury Responsiveness	0.18	0.022	0.18	8.28	0.0001
Alexa Organization	0.35	0.037	0.35	9.53	0.0001
Alexa Quality	0.007	0.020	0.07	3.58	0.0001
Alexa Download Delay	0.006	0.048	−0.11	−1.29	0.20
WebL Download Delay	0.215	0.027	−0.30	−7.97	0.0001
WebL Word Count	0.212	0.024	0.23	8.80	0.0001

Adjusted R^2 is 0.77.

one exception being Alexa download delay. The two other download delay measures (jury and WebL) do show significant association with site success. Thus, overall, the third study findings support all hypotheses.

What emerged from the third study were specific and validated dependent and independent variables and nomologically valid constructs evidenced by a strong association between constructs for Web site design and Web site success.

External Validity or Generalizability. To be useful for the research community, metrics should be applicable across other subjects and conditions (Cook and Campbell 1979). The use of varying samples of undergraduates, graduates, and executive MBA students in this series of studies suggests that the findings generalize across subjects, settings, and times (Gordon et al. 1986). The validation of metrics across these three groups provides some evidence of generalizability (Smith et al. 1996).

Discussion

Web site success is critical to the establishment of viable net-enabled business. Business-to-consumer Web sites are particularly challenging in the design process. This research identifies key metrics that can be used to identify likely elements of successful Web site design.

As Fenton (1994) counseled in his analysis of software complexity metrics, there is a challenge in the search for general measures with regard to Web site characteristics, performance, and usability metrics. However, through the use of media richness and design and usability principles, the current research has helped to define and validate measures of specific Web site attributes that are reliable, replicable, and show strong construct and nomological validity.

Findings suggest that robust metrics can be obtained from multiple sources to identify key usability elements for Web site design. Metrics from each of the three sources—jury, Alexa, and the WebL agent—were similar in their associations with Web site success. Jury data offer measures across multiple dimensions, but on the dimensions of download delay, organization, and site content, third-party and agent-gathered data appear to offer an alternative for usability and design testing.

It needs to be noted that there is no sense that any of these measures are superior to any other, but that they are all equally valid and useful. As Campbell and Fiske (1959) argue and as Boudreau et al. (2001) stress, a multimethod, multitrait interpretation of measures validates all methods symmetrically.

From a substantive point of view, site design, usability, and media richness appear to be closely associated with site success. Findings support usability principles with a set of constructs that reflect earlier work in usability (Nielsen 2000, Pearrow 2000, Shneiderman 1998) suggesting a focus on download delay, organization, and navigation. High levels of media richness are also reflected in the results, with interactivity and responsiveness constructs also being related to Web site success.

From an operational viewpoint, download delay is an easily measured element via WebL agent, Alexa, or jury. True, the agent captures only part of the total download delay time, i.e., the server-side and the transmission delay, and not the client-side delay. Nevertheless, it may be an acceptable surrogate for overall delay. Download delay is important to users and can be measured both in terms of initial access to the site

as well as within site movement, although the measures are highly correlated. These operational efficiencies are metrics that are relatively simple to measure through an agent like WebL, and they have a significant linkage with site success.

The other four factors identified in this research, navigability, content, interactivity, and responsiveness, include items that can be measured on a frequent basis through user juries. Organization of the site can be measured in terms of sequence, layout, and arrangement. Web site content includes the amount and variety of product information and can be measured by both jury information and agent word counts. Results suggest that greater information content is associated with more successful Web sites. Customization and interactivity are key Web site capabilities, and the findings suggest that successful Web sites take advantage of this through opportunities for interaction with site users. Successful sites also provide an opportunity for feedback and maintain some record of that feedback, including the availability of FAQs.

Findings suggest that Web site success can be captured through measures of frequency of use, likelihood of return, and user satisfaction. These three variables seem to compose a single construct of Web site success. Some sites might highlight likelihood of return or frequency of use, but these findings argue that these elements are reflective of the same construct (Anderson and Gerbing 1988).

In sum, the results suggest that Web site design measures demonstrate acceptable psychometric and validation properties. These metrics also appear to be important in explaining the success of Web sites. The result of this research effort is a set of parsimonious, understandable, and reliable metrics that have been tested in multiple studies, with different user groups, and across a multiyear period. Each succeeding study added components to the measures and an additional technique for data gathering. While it is likely that there will be continuing adaptation and evolution of Web site uses, technologies, and design, these metrics seem fairly stable in that the addition of data from third-party sources and agent technologies enhanced the validity without changing the fundamental constructs.

The limitations of the current research are those inherent in the development of quantitative metrics for human-computer interaction. The metrics are identifiable and replicable. However, the metrics need to be utilized across a broader sample to eventually identify the range of acceptable performance on each of the measures. It is possible that nontransaction-based Web sites and those not focused on business-to-consumer activity will be better served by additional or even other site metrics.

Implications for Web Site Managers and Designers

The metrics developed in this research can aid Web site managers and designers in creating more successful Web sites. Appropriate sequencing, layout, and arrangement of Web sites can increase navigability, and once normed through these measures, developers can test the effects of alternate site designs on success. Product information and variety of product presentations, for instance, may enhance Web site content. Designers should provide opportunities for users to customize their experience and interact with the Web site. Designers should also offer mechanisms for feedback and answering user questions. The proper balance between these elements can be explored through use of these metrics.

In testing Web sites for these elements of success, the research suggests use of a third-party-rating agent, such as Alexa, can be useful in some areas, such as organization and content. Agent-based data collection may also be helpful, as it exhibited strong correlations with other factors, and it captures key operational information, such as download delay, display rate, and word count. Use of jurors can capture these elements as well and is critical for diagnosing the interactivity and responsiveness of the Web site.

Future Research and Conclusion

Future research can address the application of these metrics to additional sites and with other juries. Metrics might be expanded to include pricing efficiencies, as well as identifying additional methods for measuring interactivity and responsiveness. Future work might also address differences in Web site task types.

Undoubtedly there will be issues keyed to specific sites and designers will test usability directly with

users. The three studies reported here represent an alternate, more generalized approach to testing usability and design. The research contributes a set of constructs and specific metric measurement techniques that can help the continuing process of improving Web site design and testing.

Acknowledgments

This research was funded in part by grants from the Honeywell Foundation, Robert H. Smith School of Business, and the University of Maryland Graduate Research Board.

Appendix. Instrument Utilized in the Web Site Analysis (Study 2 and 3)

URL: _____

Please rate the site on the following criteria using a 1 to 7 scale (1 = poor on this dimension, 7 = terrific on the dimension).

(a) Effective organization	1	2	3	4	5	6	7
(b) Provides good product information	1	2	3	4	5	6	7
(c) Presents a variety of products	1	2	3	4	5	6	7
(d) Provides information such as FAQs	1	2	3	4	5	6	7
(e) Provides feedback mechanisms	1	2	3	4	5	6	7
(f) Offers customization	1	2	3	4	5	6	7
(g) Provides significant user interaction	1	2	3	4	5	6	7

Please circle the response that best reflects your opinion of the Web site you just used.

(1) I find it easy to get this Web site to do what I want it to do.
Strongly Agree 1 2 3 4 5 6 7 Strongly Disagree
(2) I intend to browse this Web site again this semester.
Strongly Agree 1 2 3 4 5 6 7 Strongly Disagree
(3) I intend to browse this Web site frequently in the future.
Strongly Agree 1 2 3 4 5 6 7 Strongly Disagree
(4) The amount of information displayed on the screen was
Inadequate 1 2 3 4 5 6 7 Adequate
(5) The sequence of obtaining information was:
Confusing 1 2 3 4 5 6 7 Clear
(6) The information on succeeding links from the initial page was
Predictable 1 2 3 4 5 6 7 Unpredictable
(7) The Web site was:
Frustrating 1 2 3 4 5 6 7 Satisfying
(8) The layout of pages made tasks easier
Never 1 2 3 4 5 6 7 Always
(9) The speed in which the computer provided information was
Fast Enough 1 2 3 4 5 6 7 Too slow
(10) The rate at which the information was displayed was
Fast Enough 1 2 3 4 5 6 7 Too slow
(11) If you had a future need for information/service presented in this Web site, how likely is it that you would consider returning/using this site?
Very Unlikely 1 2 3 4 5 6 7 Very Likely

Alexa Ratings: Traffic: _____ Speed (1–5): _____ Quality (1–5): _____

References

Adams, D., R. Nelson, P. Todd. 1992. Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quart.* **16**(2) 227–247.

Alba, J., J. Lynch, B. Weitz, C. Janiszewski, R. Lutz, A. Sawyer, S. Wood. 1997. Interactive home shopping: Consumer, retailer, and manufacturer incentives to participate in electronic markets. *J. Marketing* **61**(3) 38–53.

Alpar, P., M. Porembski, S. Pickerodt. 2001. Measuring the efficiency of Web site traffic generation. *Internat. J. Electronic Commerce* **6**(1) 53–73.

Alwin, D. 1974. Approaches to the interpretation of relationships in the Multitrait-Multimethod Matrix. *Sociological Methodology* (1973–1974) 79–105.

Anderson, J., D. Gerbing. 1988. Structural equation modeling in practice: A review and recommended two-step approach. *Psych. Bull.* **103**(3) 411–423.

Armstrong, A., J. Hagel. 1996. The real value of online communities. *Harvard Bus. Rev.* **74**(3) 134–141.

Babbie, E. 1992. *The Practice of Social Research,* 6th ed. Wadsworth, Belmont, CA.

Bagozzi, R. 1980. *Causal Methods in Marketing*. John Wiley and Sons, New York.

——. 1981. An examination of the validity of two models of attitude. *Multivariate Behavioral Res.* **16**(July) 323–359.

——. 1983. A holistic methodology for modeling consumer response to innovation. *Oper. Res.* **31**(1) 128–176.

Barki, H., J. Hardwick. 1994. Measuring user participation, user involvement, and user attitude. *MIS Quart.* **18**(1) 59–79.

Barney, D. 2000. e-Comm intelligence report. *Network World* **17**(9) 56–57.

Berry, J. 1999. The world according to e-Biz metrics. *Internetweek* (October 4) p. 38.

Berthon, P., L. Pitt, R. Watson. 1996. Marketing communication and the World Wide Web. *Bus. Horizons* **39**(5) 24–32.

Boling, E. 1995. Usability testing for Web sites. Learning for the global community. *Seventh Annual Hypermedia Conf.* Bloomington, IN.

Bort, J. 1997. Web tracking tools chart Internet ROI. *Software Magazine* **17**(8) 65–67.

Boudreau, M., D. Gefen, D. Straub. 2001. Validation in IS research: A state-of-the-art assessment. *MIS Quart.* **25**(1) 1–24.

Buschke, L. 1997. The basics of building a great Web site. *Training & Development* **51**(7) 46–48.

Campbell, D. 1960. Recommendations for APA test standards regarding construct, trait, discriminant validity. *Amer. Psychologist* **15**(August) 546–553.

——, D. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psych. Bull.* **56**(2) 81–105.

Carlson, J., R. Zmud. 1999. Channel expansion theory and the experiential nature of media richness perceptions. *Acad. Management J.* **42**(2) 153–170.

Chain Store Age. 1997. Web-based retailers tell disparate tales at NRF. *Chain Store Age* (February) 45–52.

Churchill, G. 1979. A paradigm for developing better measures of marketing constructs. *J. Marketing Res.* **16** 64–73.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psych. Measurement* **20** 37–46.

Cohen, J. B. 1997. Too many hands in the cookie jar? *Amer. Advertising* **13**(2) 22.

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

Cronbach, L. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* **16**(3) 297–333.

Daft, R., R. Lengel. 1986. Organizational information requirements, media richness and structural design. *Management Sci.* **32** 554–571.

——, R. Lengel, L. Trevino. 1987. Message equivocality, media selection, and manager performance: Implications for information systems. *MIS Quart.* **17** 355–366.

D'Ambra, J., R. Rice. 1994. Multimethod approaches for the study of computer-mediated communication, equivocality, and media selection. *IEEE Trans. Professional Comm.* **37**(4) 231–139.

DeLone, W., E. McLean. 1992. Information systems success: The quest for the dependent variable. *Inform. Systems Res.* **3**(1) 60–95.

Dennis, A., S. Kinney. 1998. Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Inform. Systems Res.* **9**(3) 256–274.

Dickson, P. 2000. Understanding the trade winds: The global evolution of production, consumption, and the Internet. *J. Consumer Res.* **27** 115–122.

Donabedian, B., S. McKinnon, W. Bruns. 1998. Task characteristics, managerial socialization, and media selection. *Management Comm. Quart.* **11**(3) 372–400.

Ellsworth, J., M. Ellsworth. 1994. *The Internet Business Book*. John Wiley & Sons, New York, NY.

El-Shinnawy, M., M. Markus. 1998. Acceptance of communication media in organizations: Richness or features? *IEEE Trans. Professional Comm.* **41**(4) 242–253.

Evans, P., T. Wurster. 2000. *Blown to Bits*. Harvard Business School Press, Boston, MA.

Fenton, N. 1994. Software measurement—A necessary scientific basis. *IEEE Trans. Software Engrg.* **20**(3) 199–206.

Gefen, D., D. Straub, M. Boudreau. 2000. Structural equation modeling and regression: Guidelines for research practice. *Comm. AIS* **7**(7) 1–78.

Gordon, M., L. Slade, N. Schmitt. 1986. The "Science of the Sophomore" revisited: From conjecture to empiricism. *Acad. Management Rev.* **11**(1) 191–207.

Griffith, D., J. Palmer. 1999. Leveraging the Web for corporate success. *Bus. Horizons* (January-February) 3–10.

——, R. Krampf, J. Palmer. 2001. The role of interface in electronic commerce: Consumer involvement with print versus online catalogs. *Internat. J. Electronic Comm.* **5**(4) 135–153.

Ho, J. 1997. Evaluating the World Wide Web: A global study of commercial sites. *J. Comput.Mediated Comm.* **3**(1) ⟨www.ascusc.org/jcmc/vol3/issue1/ho.html⟩.

Hoffman, D., T. Novak. 1996. Marketing in hypermedia computer-mediated environments: Conceptual foundations. *J. Marketing*, **60**(3) 50–68.

——, ——, P. Chatterjee. 1995. Commercial scenarios for the Web: Opportunities and challenges. *J. Comput. Mediated Comm.* **1**(3) ⟨www.ascusc.org/jcmc/vol1/issue3/hoffman.html⟩.

Huang, W., R. Watson, K. K. Wei. 1998. Can a lean e-mail medium be used for rich communication? A psychological perspective. *Euro. J. Inform. Systems* **7**(4) 269–274.

Jarvenpaa, S., P. Todd. 1997. Consumer reactions to electronic shopping on the World Wide Web. *Internat. J. Electronic Commerce* **1**(2) 59–88.

Johnson, R., D. Wichern. 1998. *Applied Multivariate Statistical Analysis*. Prentice Hall, New York.

Landis, J., G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* **22** 79–94.

Lederer, A., D. Maupin, M. Sena, Y. Zhuang. 1998. The role of ease of use, usefulness and attitude in the predication of World Wide Web usage. *Proc. 1998 ACM SIGCPR Conf.*, March 26–28, Boston, MA.

Levi, M., F. Conrad. 1996. A heuristic evaluation of a World Wide Web prototype. *Interactions* **3**(4) 1–12.

Levine, R. 1996. Guide to Web style. ⟨http://Web.sun.com/styleguide/tables/welcome.html⟩.

Machlis, S. 1998a. Quick study: Cookies are a marketer's dream, but do they watch too closely? *Computerworld* **32**(28) 25.

——. 1998b. Site redesigns keep it simple. *Computerworld* **32**(43) 43–44.

Markus, M. 1994. Electronic mail as the medium of managerial choice. *Organ. Sci.* **5** 502–527.

McQuarrie, E., J. Munson. 1991. A revised product involvement inventory: Improved usability and validity. John F. Sherry and Brian Sternthal, eds. *Advances in Consumer Research*, vol. 19. The Association for Consumer Research, Provo, UT.

Messmer, E. 1999. Internet holiday shopping stresses online toy stores. *Network World* **16**(49) 6.

Munch, J., G. Boller, J. Swasy. 1993. The effects of argument structure and affective tagging on product attitude formation. *J. Consumer Res.* **20**(2) 294–302.

Neely, A. 1998. *Measuring Business Performance*. Profile Books, Ltd., London, U.K.

Niccolai, J. 1997. Web-analysis tools can measure more than a site's success. *InfoWorld* **19**(14) 62.

Nielsen, J. 1993. *Usability Engineering*. Morgan Kaufmann, New York.

——. 2000. *Designing Web Usability*. New Riders Publishing, Indianapolis, IN.

Nunnally, J. 1967. *Psychometric Theory*. McGraw-Hill, New York.

——. 1978. *Psychometric Theory*, 2nd ed. McGraw-Hill, New York.

Palmer, J. 1997. Retailing on the WWW: The use of electronic product catalogs. *Internat. J. Electronic Markets* **7**(3) 6–9.

——, D. Griffith. 1998. An emerging Web site design model for marketing. *Comm. ACM* **41** (3) 44–51.

Pearrow, M. 2000. *Web Site Usability*. Charles River Media, Rockland, MA.

Petty, R., J. Cacioppo. 1979. Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *J. Personality and Soc. Psych.* **37**(9) 1915–1926.

Picarille, L. 1997. PointCast proposes Web-hit yardstick. *Comput. Reseller News* **757** p. 85.

Pitkow, J., C. Kehoe. 1996. Emerging trends in the WWW user population. *Comm. ACM* **39**(6) 106–108.

Price, M. 1997. What makes users revisit a Web site? *Marketing News* (March 17) p. 12.

Przystupa, F. 1993. Interactive interfaces between collective support systems and their users. R. Glanville and G. de Zeeuw, eds. *Interactive Interfaces and Human Networks*. Thesis publishers, Amsterdam, North Holland, 85–96.

Radosevich, L. 1997. Fixing Web-site usability. *InfoWorld* **19**(50) 81–82.

Randall, D. 1997. Consumer strategies for the Internet: Four scenarios. *Long Range Planning* **30**(2) 157–168.

Rasmussen, J. 1996. Design with the user in mind, and you can't go awry. *Comput. Canada* **22**(25) 47–48.

Rice, R. 1992. Task analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. *Organ. Sci.* **3**(4) 475–500.

——. 1993. Media appropriateness: Using social presence theory to compare traditional and new organizational media. *Human Comm. Res.* **19** 451–484.

——, J. D'Ambra. 1998. Cross-cultural comparison of organizational media evaluation and choice. *J. Comm.* **48**(3) 3–20.

——, D. Shook. 1990. Relationships of job categories and organizational levels to use of communication channels, including electronic mail: A meta-analysis and extension. *The J. Management Stud.* **27**(2) 195–229.

——, D. Hughes, G. Love. 1989. Usage and outcomes of electronic messaging at an R&D organization: Situational constraints, job level, and media awareness. *Inform. Tech. & People* **5**(2) 141–161.

Rose, G., D. Straub. 2001. The effect of download time on consumer attitude toward the e-Service retailer. *e-Service J.* **1**(1) 55–76.

——, H. Khoo, D. Straub. 1999. Current technological impediments to business-to-consumer electronic commerce. *Comm. AIS* **1** (16) 1–74.

Ross, D. 2001. Electronic communications: Do cultural dimensions matter? *Amer. Bus. Rev.* **19**(2) 75–81.

Schmitz, J., J. Fulk. 1991. Organizational colleagues, media richness, and electronic mail—A test of the social influence model of technology use. *Comm. Res.* **18** 487–523.

Schonberg, T., R. Hoch, M. Podlaseck, S. Spraragen. 2000. Measuring success. *Comm. ACM* **43**(8) 53–57.

Schubert, P., D. Selz. 1998. Web assessment: A model for the evaluation and the assessment of successful electronic commerce applications. *Proc. 31st Hawaii Internat. Conf. Systems Sci.*, Hawaii.

Selz, D., P. Schubert. 1997. Web assessment: Some evidence from successful interactive WWW applications. *Internat. J. Electronic Markets* **7**(3) 1–4.

Shapiro, C., H. Varian. 1999. *Information Rules*. Harvard Business School Press, Boston, MA.

Shneiderman, B. 1998. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Reading, MA.

Smith, H., S. Milberg, S. Burke. 1996. Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quart.* **20**(2) 167–196.

Spool, J. 1997. *Web Site Usability: A Designers Guide*. User Interface Engineering (UIE), San Francisco, CA.

Stone, M.L. 1999. Server logs help shape Web strategies. *Advertising Age's Bus. Marketing* **84**(1) 19–20.

Straub, D. 1989. Validating instruments in MIS research. *MIS Quart.* **13**(2) 147–169.

——, E. Karahanna. 1997. Knowledge worker communications and recipient availability: Toward a task closure explanation of media choice. *Organ. Sci.* **9**(2) 160–175.

——, R. Watson. 2001. Research commentary: Transformational issues in researching IS and net-enabled organizations. *Inform. Systems Res.* **12**(4) 337–345.

——, M. Limayem, E. Karahanna. 1995. Measuring system usage: Implications for IS theory testing. *Management Sci.* **41**(8) 1328–1342.

Suh, K. 1999. Impact of communication medium on task performance and satisfaction: An examination of media-richness theory. *Inform. & Management* **35**(8) 295–312.

Took, R. 1990. Putting design into practice: Formal specification and the user interface. M. Harrison, H. Thimbleby, eds. *Formal Methods in Human-Computer Interaction*. Cambridge University Press, Cambridge, U.K., 82–96.

Trevino, L., R. Lengel, R. Daft. 1987. Media symbolism, media richness, and media choice in organizations: A symbolic interactionist perspective. *Comm. Res.* **14**(5) 553–574.

——, ——, W. Bodensteiner, E. Gerloff, N. Muir. 1990. The richness imperative and cognitive style. *Management Comm. Quart.* **4** 176–197.

Umesh, I., R. Peterson, M. Sauber. 1989. Interrespondent agreement and the maximum value of Kappa. *Educational and Psych. Measurement* **49** 835–850.

Wilson, T. 1999. Service keeps tabs on partner sites. *Internetweek* (September 13) p. 780.

Detmar Straub, Senior Editor. This paper was received on January 10, 2000, and was with the author 10 months for 3 revisions.