WebQual: An Instrument for Consumer Evaluation of Web Sites

Eleanor T. Loiacono, Richard T. Watson, and Dale L. Goodhue

ABSTRACT: Despite the critical need to know how consumers' perceptions of Web sites influence their behavior, and especially their intention to revisit or purchase, there is no extant general measure for evaluating Web sites and no consensus on what such an instrument should measure. The authors used the Theory of Reasoned Action and the Technology Acceptance Model to develop the WebQual instrument for consumer evaluation of Web sites. They refined it through a literature review and interviews with Web designers and users, and tested it using four samples of Web consumers. WebQual includes 12 dimensions (informational fit-to-task, tailored information, trust, response time, ease of understanding, intuitive operations, visual appeal, innovativeness, emotional appeal, consistent image, on-line completeness, relative advantage) and shows strong measurement validity. It is a highly validated instrument that can provide both wide- and fine-grained measurements of organizational Web sites.

KEY WORDS AND PHRASES: Consumer evaluation, instrument development, TAM, Technology Acceptance Model, Theory of Reasoned Action, TRA, Web site evaluation.

Web sites are a critical component of the rapidly growing phenomenon of e-commerce. Worldwide, business-to-consumer (B2C) retail sales netted $172 billion in 2005. Internet retail sales are expected to grow to $329 billion by 2010 [54]. Web sites play a significant role in the overall marketing communication mix [14]—they complement direct selling activities, provide supplemental material and basic company information to customers, and project a corporate image. Businesses are eager to develop means for measuring and analyzing consumer perceptions of Web site designs.

Many academic researchers studying e-commerce have been forced to develop their own Web evaluation instruments [3, 29, 58, 61, 65, 72]. Given the amount of practitioner interest and research in this area, a well-validated, generally available instrument would be very valuable. It is time for the IS field to develop such an instrument, following a rigorous and comprehensive development method. The evolving history and science of instrument development and testing in behavioral research [7, 8, 20], and especially in IS research [26, 42, 53, 92], suggest that the development of an instrument to capture consumers' evaluations of a Web site should at a minimum have the following characteristics:

1. A careful review of the literature and relevant practice to identify the specific constructs that should be part of the measure. Too rapid a convergence on a particular set or framework of constructs runs the risk of excluding important characteristics of the phenomena. Out of this comes a set of clearly defined and distinct constructs.
2. Multiple questions for each construct, with each construct being unidimensional. Since error is common to all measures, multiple

        questions make it possible to assess measurement quality and to address measurement problems where they exist.
3. Adequate discriminant validity tests as well as reliability tests. Reliability without discriminant validity can mislead developers into thinking that the constructs are distinct when, in fact, they are not.
4. Samples of a size appropriate to the statistical techniques used for assessing measurement validity.
5. A test of nomological or predictive validity. It is necessary to know that the measures of the constructs interact in reasonable ways with other constructs.

If consumer evaluation of Web sites is the target of the measurement effort, two additional requirements must be met:

6. The focus must be on the consumer's reactions to many specific Web sites, and not on e-commerce or Web sites in general. An instrument intended for this purpose must be tested in this context.
7. The scope of data collection must include reactions to many different Web sites (even if only one per respondent). Given the method of testing for discriminant validity, restricting the focus to a single Web site (even if evaluated by many consumers) will surface individual differences of opinion about a single object, rather than the ability of the instrument to accurately track the dimensions along which consumers perceive Web sites as differing.

An instrument whose development lacks any of the seven characteristics outlined above should not be seen as a validated general-use instrument for consumer evaluation of Web sites. While for any specific development effort there is room for different judgments about whether the first characteristic has been appropriately met, determining compliance for the other conditions is fairly straightforward. Some research on consumer evaluations of Web sites has been published (e.g., [11, 84, 111]), but an extensive literature review found no study that met all the conditions described above.

More specifically, Schubert and Selz developed their WA—Web Assessment tool (www.businessmedia.org/businessmedia/businessmedia.nsf/pages/wa_tool.html) based on the three phases of a transaction (information, agreement, and settlement) augmented by a community phase [85]. Subsequent research extended the model to create an Extended Web Assessment Model (EWAM) by including elements of TAM, social influence, and a review of four practitioner reports on Web evaluation [84]. The models were tested with samples of 55 and 20 respondents, respectively, not enough for a careful assessment of measurement validity.

In another customer-centric endeavor, Barnes and Vidgen refined a measure of Web site quality over four versions [11].[1] Their WQ is based on quality function deployment [15]. Six graduate students generated the initial items, too narrow a base for establishing content validity. The samples for the first three versions were small (46, 54, and 39 respondents), and the factor struc-

ture varied significantly across the four versions, raising questions about the underlying stability of the initial conceptualization.

As a third example, Yoo and Donthu developed SITEQUAL [111] by asking students in two marketing classes to generate appropriate questions. Fifty-four unique items were generated, and they were the basis for an instrument completed by 69 students for three self-selected sites. Although the sample was small, exploratory factor analysis (EFA) was used to reduce the instrument to 38 items and nine factors of two broad sets: vendor-related and site quality. The first set of factors was removed because the researchers wanted to focus on site quality. An 18-item instrument measured the remaining four factors (ease of use, design, processing speed, and security). Confirmatory factor analysis (CFA), apparently using the same data, indicated a poor fit and the model was respecified. After several iterations, the instrument was reduced to nine items to measure the four factors (only two items for most factors). A validation study with 47 subjects, each of whom evaluated four sites ($n = 187$), resulted in reliabilities in the range 0.69 to 0.83 and good fit indices. In the authors' view, SITEQUAL's original set of items was too narrowly based, and most of its final factors were measured by only two items.

It is also instructive to look at the articles in the *Information Systems Research* (*ISE*) Special Issue on Measuring e-Commerce Net-Enabled Organizations (June and September 2002). Six of the articles in this double issue include consumer evaluations of Web sites. None of the six articles meets all seven conditions outlined above.[2] Relative to condition 1 (coverage of appropriate constructs), a comparison of the constructs targeted in the six measures demonstrates that there was no convergence on what dimensions should be measured, as shown in Appendix 1. Moreover, as also shown in Appendix 1, none of the measures in the special issue of *ISR* has the breadth and coverage of the measure described in this paper.

In fairness, the authors of the *ISR* articles had other research agendas, and none of them took as a primary purpose the development of a general instrument for consumer evaluation of Web sites. The researchers who authored the articles all needed a measure for evaluating Web sites, but they developed different measures. Although valuable in their own ways, none of these measures has the focus and rigor that would be required to win wide acceptance for a measure of a construct so critical in research and practice. If a carefully developed and validated general measure had been available, these researchers (and others) would have been able to focus more of their research time and energy on their research agendas rather than on measurement development.

The present article addresses the need for a general measure of consumer evaluations of business-to-consumer Web sites. The instrument it describes, WebQual, has been reported in various conference papers, but there has been no published description of its careful development through multiple iterations of data collection, measurement validity analysis, and instrument refinement. Nevertheless, as evidence of the demand for such an instrument, there have been more than 150 requests for information about WebQual from researchers and companies, and it has already been cited more than 15 times in conferences, workshops, and journals.

One cannot evaluate an object without some context. Of particular concern to designers of Web sites is the question of whether consumers' reactions to a Web site are likely to induce subsequent visits or purchases from the site in the future [38]. Thus the conceptual context used in developing WebQual was: What perceived characteristics of a Web site will affect a consumer's decision to reuse the site?

The IS and marketing literature were drawn upon to identify the specific beliefs important to predicting consumer reuse of a Web site. In addition, exploratory research was conducted using expert judges. An initial instrument was refined through three different versions. At each step, the instrument's measurement validity was carefully analyzed, using large samples (510, 336, 311, and 377 students) to refine the conceptualization and the questions. The final version contains 36 questions on 12 characteristics of a Web site. It demonstrates strong measurement validity, and it predicts intention to buy from or revisit a Web site.

Research Framework

Theoretical Underpinnings

Using a Web site can be both an IS interaction and a marketing interaction [38]. To use a Web site, one must employ computer hardware and software focused on information storage, display, processing, or transfer. In addition, many aspects relate to marketing interactions. Brand and company images affect perceptions, information is passed, consumer questions are answered, and purchases are made. Therefore both IS and marketing literature were employed in conceptualizing the way consumers evaluate a Web site.

The Theory of Reasoned Action (TRA), which is used extensively in marketing and IS research, argues that individuals evaluate the consequences of a particular behavior and create intentions to act that are consistent with their evaluations (*see top of Figure 1*) [4, 35]. Davis applied TRA to a class of behaviors that can be loosely defined as "using computer technologies," and produced the Technology Acceptance Model (TAM), one of the most widely cited pieces of IS research [25, 103]. Davis argues that two beliefs are predominant in predicting the behavior of using computer technologies: perceived ease of use and perceived usefulness. Through an extensive stream of research, Davis and others have developed strong measures of these two beliefs and demonstrated their predictive power in a number of contexts, including use of word processors, e-mail, drawing tools, and hospital information systems [2, 26, 97].

Davis found that attitudes did not completely mediate the relationship between beliefs and intentions (as Fishbein had suggested), and argued that it made more sense to focus on measuring these beliefs as direct predictors of intentions rather than trying to measure attitudes as well. Davis also initially argued that in the context of using computer technologies (at least in the domains he studied), subjective norms did not seem to be a significant predictor of intentions.
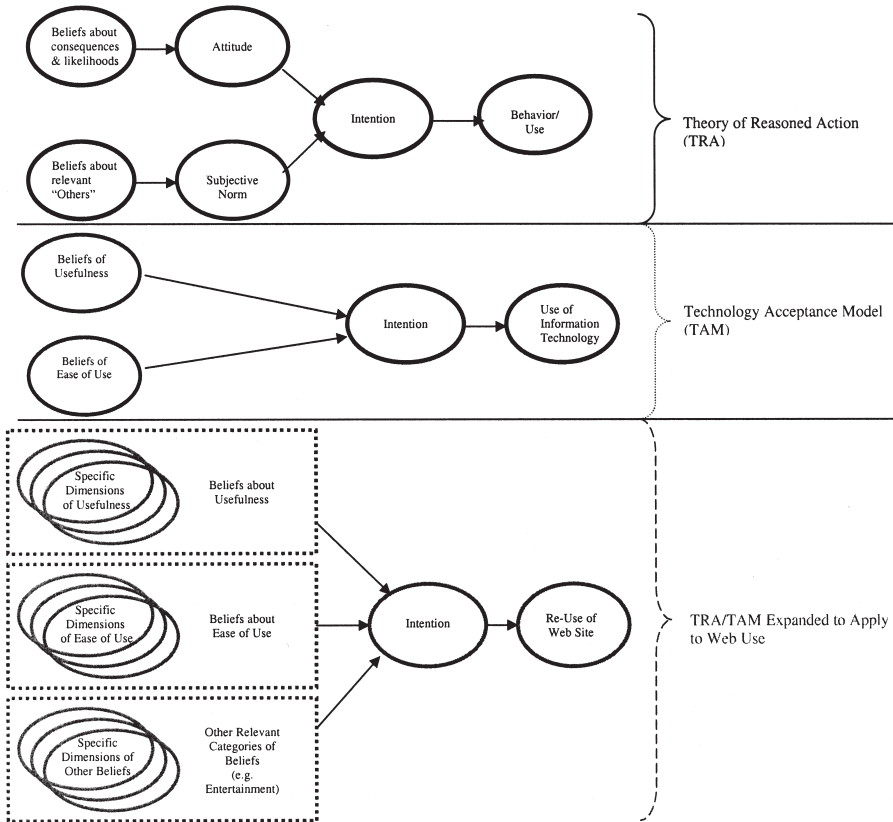
Figure 1. Research Framework

There is some disagreement about ignoring social norms in predicting IS use (e.g., [19]), but several studies have successfully excluded social norms, focusing only on ease of use, usefulness, trust, and so forth, as predictors of e-commerce use [52, 74, 95]. Since the interest here is in whether an individual will use a specific Web site, there is no reason why subjective norms should have a large impact on behavior, as opposed to other behaviors where norms might have more of an impact (e.g., smoking or dieting). For the most part, use of a particular site is a private affair, not visible to one's peers. Peer pressure might encourage an individual to use the Web in general, but not to revisit a particular Web site. Given this argument, Davis's findings, and the need to contain the scope of the research to a reasonable size, social norms have also been excluded from this study.

TAM only identifies two very general beliefs: ease of use and usefulness (*see middle of Figure 1*). The effort to develop an instrument for evaluating Web sites begins by looking both *beyond* and *within* ease of use and usefulness. There are two reasons why the focus is not narrowed to only ease of use and usefulness. First, evidence suggests that Web use is driven by some additional factors beyond these two. In particular, Web use may have entertainment

value that is not easily captured by ease of use or usefulness [50, 86]. For this reason it is important to consider the possibility of adding to Davis's two general constructs.

Second, following Goodhue and Thompson, an instrument for evaluating a Web site will be most useful to businesses if it can identify in more detail the specific aspects that make the Web site easy to use or useful to consumers [43]. The greater clarity of detail is important conceptually because some aspects may be found empirically to be more important than others in determining consumer behavior. This is significant from a practical business standpoint because without a finer-grained measure than ease of use or usefulness, a business might not know what changes to make in a Web site rated low in usefulness. It is necessary, therefore, to consider the possibility of multiple distinct dimensions of ease of use and of usefulness, as well as other categories of beliefs, such as entertainment, which together predict intention to reuse a Web site (*see bottom of Figure 1*).

Method

There are many frameworks for thinking about measurement validity. The frameworks proffered by Bagozzi and by Bagozzi and Phillips are used in this paper because of their comprehensive coverage of the six key components of validity shown in Table 1, which are explained along with the instrument development process [7, 9].

Instrument Development Process

The overall process of instrument development took place in four stages that together address all of Bagozzi's validity concerns. The process also meets the seven requirements stated at the beginning of this paper for a general-purpose measure of consumer evaluations of Web sites.

1. *Defining the Dimensions.* Various techniques were used to determine and define the categories of beliefs that should be considered: a literature review, exploratory surveys, and expert judges. Both theory and an understanding and interpretation of the phenomenon laid the work's foundation. (Requirement 1)
2. *Developing the Items.* Multiple questions were developed for each of the dimensions identified in Stage 1. The initial result was an 88-item instrument that measured 13 distinct beliefs about a specific Web site. (Requirements 2 and 6)
3. *Refining the Instrument.* The instrument was refined by administration to two different samples ($N = 510$ and $N = 336$), each respondent rating one of 12 different sites. After each administration, the measurement validity of the constructs was analyzed, problem questions were pruned, revised, or replaced, and redundant dimen-

Validity issue	Concern
Theoretical meaningfulness of concept	Constructs well defined
	Making theoretical sense
Observational meaningfulness of concept (content validity)	Measures correspond to theoretical constructs
Internal consistency	Maximally similar measures of same construct agree (i.e., reliability)
Discriminant validity	Distinct constructs can be distinguished
Convergent validity	Maximally dissimilar measures of same construct correlate (e.g., do a collection of questions on a questionnaire correlate with an overview question or with some objective measure)
Nomological validity	Making sense in the larger theoretical framework

Table 1. Validity Concerns.

Sources: Based on Bagozzi [7] and Bagozzi and Phillips [9].

      sions were collapsed. This resulted in an instrument with 36 questions measuring 12 dimensions. (Requirement 7)

4. *Confirmatory Assessment of Validity.* A confirmatory analysis of the overall measurement validity of the final instrument was conducted using two new samples, one of 311 students collected in the summer and fall of 2000, and another of 377 collected in 2002/2003. In both samples, the instrument demonstrated strong measurement validity for the four validity issues that can be empirically assessed—that is, Bagozzi's validity concerns (*see Table 1*). (Requirements 3, 4, and 5)

Stage 1: Defining the Dimensions

A four-pronged effort was employed more or less simultaneously to determine the pertinent dimensions of Web site evaluation and establish content validity. First, a review of the IS and marketing literature revealed existing constructs related to Web design and customer satisfaction. Popular press publications were also examined to ensure that no factor was overlooked because of the "newness" of the Web. In parallel with this effort, three exploratory research projects ensured the comprehensiveness of the constructs relative to the domain of the Web: soliciting criteria from Web surfers, interviewing professional Web designers, and studying a large organization's standards for Web site design.

    Two theoretical perspectives guided the effort to identify all the relevant distinct constructs that might affect a consumer's intent to reuse a Web site. The first was the Technology Acceptance Model [25, 26]. As explained earlier, an effort was made to look both within and beyond ease of use and usefulness.

    The usefulness realm was explored more deeply by utilizing the insight from task-technology fit that a technology is useful when it fits the task [43]. This made it necessary to identify the tasks a Web site consumer might engage

in and aspects of the technology that either support or thwart the consumer's accomplishment of these tasks. Two generic consumer Web tasks were identified: *gathering information* (about a company, a product, or some nonbusiness issue) and *carrying out a transaction* (related to a standard purchase or to some more complex interaction, such as a service).

Finally, it is recognized that use of a Web site sometimes goes beyond utilitarian aspects (i.e., usefulness) to include *entertainment value* [14, 76, 78]. Visitors may be interested in several of these categories at the same time (e.g., entertainment and gathering information), but conceptually they can be focused on separately in seeking to understand the way a Web site affects a consumer engaged in that task.

Therefore the search for distinct dimensions for evaluating Web sites begins with a framework of four categories: ease of use, usefulness in gathering information, usefulness in carrying out transactions, and entertainment value.

- *Dimensions Relating to Ease of Use.* The literature covers not only the traditional idea of information being easy to read and understand [25, 34, 48, 60, 96] but also the importance of ease of operation (*intuitive operations*) and navigation (*ease of understanding*) for a Web site [13, 68].
- *Dimensions Relating to Usefulness in Gathering Information.* Regardless of the information search activity, certain characteristics of the process emerge as important. These include *information quality* [10, 12, 27, 55, 93, 99, 107, 108] and the ability of the site to provide the consumer with *tailored information* through such things as search fields [41, 91].
- *Dimensions Relating to Usefulness in Carrying Out Transactions.* A Web site must support the consumer's desire to carry out a transaction on-line. The site must meet the consumer's functional task needs (*functional fit-to-task*) [43, 94] and expected response times (*response time)* [64, 87], and gain the consumer's *trust* [30, 46, 49, 51, 67]. The *customer service* provided by a firm enhances or detracts from the consumer's ability to complete the task [44, 56, 57, 73, 110]. Other important factors are the ability to complete necessary transactions on-line (*on-line completeness*) [87], the *relative advantage* to consumers of completing transactions over the Web, as compared to alternative means [68, 82, 87], and the projection of a *consistent company image* across all points of contact with the customer [64, 80, 87, 109].
- *Dimensions Related to Entertainment Value.* Finally, some consumers seek a "full experience"—they want to be entertained by the process of searching. Whether they make a purchase or not, they simply enjoy "strolling down the aisles" and want to be entertained along the way. For these consumers, the Web site must create a pleasant "experience." It must be *visually appealing* [34, 39, 47], have an *innovative* flare separating it from just any old site [1, 32, 33], and be *emotionally appealing* [28, 81] so as to encourage continued browsing [23, 40, 70, 102, 103, 104]. See [63] for details on all these dimensions.

The literature search identified 14 constructs. Related to ease of use were ease of understanding and intuitive operations. Related to gathering information were information quality and tailored information. Related to carrying out transactions were functional fit-to-task, trust, response time, consistent image, on-line completeness, relative advantage, and customer service. Related to entertainment, visual appeal, and innovativeness was emotional appeal. For comparison and contrast, Appendix 1 shows the constructs measured by Web-Qual and by five articles measuring e-commerce [29, 58, 61, 65, 72]. WebQual surpasses the other measures in terms of coverage and completeness.

Exploratory Research

In parallel with the literature review, three exploratory research projects were conducted to ensure that the model generated was comprehensive and that items were drawn both from the literature and consumers. This approach ensured that the study did not miss any key aspect that had not yet made it into the established literature.

First, four groups of about 20 students each (approximately 80 students total) were asked to divulge their individual ranking criteria for high- versus low-quality Web sites. Each group ranked 10 Web sites within one of four product/service categories: CDs, books, hotel reservations, and airline reservations (40 Web sites in all). Of particular interest for this project, the subjects also recorded the factors that in their view differentiated the sites. Two judges categorized the comments into higher-level categories.

Second, mirroring the development of SERVQUAL, practitioner knowledge was also tapped to ensure that no key factors were overlooked [73]. Companies are concerned with what their customers want in a Web site and have developed criteria based on these desires. Using telephone interviews, 10 Web designers were asked to explain the criteria they used in creating Web sites.

Third, along the same lines, the Web design criteria used by a Fortune 500 company were studied.

All this exploratory research lent support to nine of the WebQual dimensions uncovered in the IS and marketing literature: tailored information, response time, ease of understanding, intuitive operations, visual appeal, emotional appeal, consistent image, on-line completeness, and relative advantage. Overlap on dimensions arising out of the literature and the exploratory research was a welcome sign that the research effort was, in fact, comprehensive, and had not missed any critical aspects of Web site evaluation.

Finally, a review of research performed at the same time or subsequent to the initial work on WebQual reveals that other researchers concur with the importance of particular dimensions identified above for WebQual, but none of these researchers has provided as complete a picture as WebQual. This further confirms the relevance of the total set of dimensions for Web site evaluation: information quality [18], functional fit-to-task [79], tailored information [83], trust [62, 66, 75, 90, 98], response time [36], ease of use [69, 88], intuitive operations [69, 88], visual appeal [45], consistent image [18, 90], relative advantage [18, 77], and customer service [83, 112].

Underlying Structure of Web Site Evaluation

Each of the 14 dimensions is assumed to be a distinct construct capable of varying independently from the others. For example, a Web site might have high ease of understanding on each page, but not have intuitive operations, thus making navigation more difficult. This is a total disaggregation model [8]. Further, the overall structure of a Web evaluation is conceptualized as bottom up, meaning that the 14 constructs are not treated as "reflections" of some single underlying overall construct, but instead the overall evaluation is seen as "produced by" the combination of the 14 underlying constructs [8]. This conceptualization of the structure requires a demonstration that the measures of each construct are distinct (discriminant validity) even if they are correlated.

Companies conceivably could attempt to address inadequacies in any one of the 14 constructs independently of the others. However, the link between aspects of Web site design and consumer perceptions might be more straightforward on some dimensions (say, response time) than on others (say, visual appeal or emotional appeal).

Narrowing the Scope of the Current Research

All but one of the identified dimensions for evaluating a Web site can be assessed after a single visit to a Web site. Only customer service suggests the need for multiple interactions before an assessment can be made. In fact, customer service includes both one-time interactions with a company via the Web site (e.g., obtaining company information) and multiple interactions (e.g., receiving an e-mail response following an inquiry). In order to restrict the scope of the current research to a manageable size, customer service measures requiring multiple interactions were not included in the initial instrument. Subsequent research could then focus on refining the dimensions and developing measures of those aspects of customer service not assessable after a single visit.

Stage 2.0: Developing the Items

The second of Bagozzi's validity concerns, observational meaningfulness (*see Table 1*), refers to the extent to which the questions (i.e., the operationalizations) actually cover all relevant aspects of the concept (content validity) and whether there is a persuasive reason to believe there is a link between the questions and the underlying constructs they are intended to measure. Scale development can be either deductive or inductive [50]. Both approaches were incorporated through an extensive literature review (inductive) and an exploratory research phase (deductive), as previously described.

An initial set of 142 candidate items for the 13 dimensions was developed based on the literature, the exploratory research with Web shoppers, and the discussions with Web developers. This was done by carefully reviewing the construct definitions, any items for similar constructs from published research,

and the comments made and language used by subjects in the first round of exploratory research. With these sources in mind, appropriate items were developed to match the construct's definition, with the goal of making items clear, straightforward, and in the everyday language of consumers.

This set of 142 items was then reduced, using an approach similar to the one used by Davis in his pretest of measures for the TAM [25]. Twenty Web consumers from a large southeastern university (5 graduate students and 15 undergraduates) rated the items on how well they corresponded to descriptions of the four higher-level dimensions of a Web site (ease of use, usefulness, entertainment, and complementary relationships).[3] (The rankings were done at an individual level with no collaboration between consumers.) Items that ranked lowest were discarded, keeping at least three items for each of the original 13 constructs. This resulted in an initial instrument of 88 items covering 13 constructs.

Stage 2.1: Experimental Procedure for Rounds 1–4

The students who participated in the study received extra course credit. Those who preferred not to participate could receive the same amount of extra credit by writing a short report on the Web. Many students opted to do neither, indicating that they did not feel coerced to participate. The participation process, and all the other elements of the study, was approved by the university's human subjects committee.

In every round, the subjects were given a context. For example, they might be told "Imagine it is your friend's birthday, and you are searching for a good gift—a book." They would then be told to explore a designated Web site for at least 10 minutes as if they were considering which book to buy for their friend, and after this to complete the questionnaire. In rounds 1 and 2, they completed the survey individually in a lab environment. The administrator of the survey reviewed the directions with each group—explaining that they were to indicate their level of agreement for each statement by circling the appropriate number between 1 and 7 (strongly disagree to strongly agree) and reminding them that this was to be done individually. In rounds 3 and 4, subjects were given written instructions and allowed to complete the survey individually on their own time.

Twelve different Web sites were used, three each of four different types: books, CDs, hotels, and airlines (*see Table 2*). Both service and product sites were chosen to ensure that there was no bias to either type of site. CD and book sites are obvious categories for students. For this student population (a fairly affluent group), airline and hotel sites were also appropriate. Many of the students travel via airplane to their homes, and others indicated that they used the Web to plan spring break trips. Any students who had not actually used the Web to make hotel or airline reservations certainly knew others who had and could easily identify with the task. Since the object of interest included not only consumers who had made a Web purchase, but those who were potential Web purchasers, it was appropriate to include all four types of Web sites. The 12 specific sites were chosen for their variability, based on

	Site 1	Site 2	Site 3
Products			
CDs	Music Point	Emusic	CDNow
Books	Amazon	ReadersNDEX	Waterstones
Services			
Airline reservations	4 Airlines	Period.com	Strangeways
Hotel reservations	First option	Hotels & Travel on Net	Places to stay

Table 2. Web Sites for Data Collection.

rankings of specific Web sites generated by subjects in the earlier exploratory research phase.

Stage 3: Refining the Instrument

Round 1

Item-order bias was prevented by creating two random-order versions of the initial 88-item instrument. To prevent inflating reliabilities from artificially high correlations where subjects answered adjacent questions using anchoring and adjustment, multiple items pertaining to a single construct were separated from each other. Items were measured on a seven-point Likert scale. Reverse-scored items were included to keep respondents alert while completing the survey and to eliminate response bias [49].

Since there were 88 questions in the initial questionnaire, and the rule of thumb for factor analysis calls for at least five times as many observations as there are variables to be analyzed, at least 440 subjects were required [48]. In round 1, data were collected from 510 undergraduates enrolled in an introductory IS course. The specific characteristics of each round of data collection are listed in Table 3. The ages and genders of the subjects matched closely with those of the largest group of Internet users—18 to 29 years old, approximately 50 percent female and 50 percent male [16, 24].

The data analysis and purification phase of the study was guided by the recommended procedure for instrumental development [25, 26]. The first pass of data analysis and purification was intended to improve measurement validity by identifying and removing the worst-performing items. First, items that reduced the reliability (Cronbach's alpha) of any of the 13 constructs were deleted and the process continued until no item's removal improved a construct's overall alpha. The end-result was the removal of 11 items. Second, items found to possess low correlations with similar traits (i.e., less than 0.40) were removed from the instrument. This followed the modified multitrait-multimethod process of item deletion as described by Goodhue [17, 42]. A total of 13 items were deleted during this phase. The final step consisted of removing items that appeared to have serious discriminant validity problems. Items were removed if they correlated much more highly with items measur-

	WebQual round 2 spring 2000	WebQual round 3 summer/ fall 2000	WebQual round 4 * 2002/2003	Recommended cutoff
Chi-square	1059.50	1095.75	1085.70	Not applicable in
(*p*-value)	(0.000)	(0.000)	(0.000)	large models, large
df	df = 528	df = 528	df = 440	sample sizes
RMSEA	0.055	0.059	0.062	< 0.06 –0.08
SRMR	0.047	0.053	0.049	< 0.06 –0.08
RNI	0.95	0.92	0.92	≥ 0.90
NNFI	0.94	0.91	0.90	≥ 0.90

Table 3. Overall Fit of Full WebQual Model.

Notes: df = degrees of freedom, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual, RNI = relative noncentrality index (equivalent to CFI), NNFI = non-normed fit index (equivalent to TLI). * Only 11 constructs were included in Round 4. Emotions were not included.

ing different constructs than with items in their intended construct [17, 42]. Under these criteria, eight items were deleted.

Following Straub, no effort was made to remove all violations of Campbell's discriminant validity rule, since spurious correlations can occur by chance in such a large correlation matrix [92]. Further, as planned, more precise means of detecting discriminant validity problems would be employed in future rounds. In particular, an effort was made to note possible discriminant validity problems between information quality and functional fit-to-task. Given the strong conceptual arguments for the distinctness of these constructs, they were not collapsed at this time, and the second round would show whether the problems continued.

After items were deleted, each construct was reviewed to ensure that at least five items per construct remained. (If measurement validity problems were discovered in future founds, this would make it possible to drop up to two items for each construct and still have at least three items per construct.) For constructs that were underrepresented, additional items closely related to the remaining items were generated by the researchers,[4] keeping in mind the conceptual definitions of the constructs and the language used by consumers in the exploratory studies. Twenty-seven items were added, resulting in an 83-item instrument.

Round 2

In round 2, data were collected from 336 undergraduate students (*see Table 4*). A two-step process was employed to select the questions to be included in the next version and make an initial test of their validity.

Selection of Best Questions for Each Dimension. Since a construct should have at least three items, and lengthy questionnaires typically have a lower response rate, it was decided to include only the top-three items loading on each construct in the final questionnaire [6, 22]. An EFA of all 83 items was

	Round 1 winter 2000	Round 2 spring 2000	Round 3 summer/ fall 2000	Round 4 2002/ 2003
n	510	336	311	377
Number of items	88	83	36	36
Average age	20	21	20	20
Gender	Male = 50% Female = 50%	Male = 62 % Female = 38 %	Male = 52% Female = 48%	Male = 51% Female = 49%
Ever made a Web purchase	Yes = 54% No = 46%	Yes = 67% No = 33%	Yes = 64% No = 36%	Yes = 73% No = 27%
Average purchases in past 30 days	1.47	1.71	.81	1.11
Average number of years using Web	4.02	4.66	4.75	6.21
First time on Web site	Yes = 85% No = 15%			
Heard of company before		Yes = 18% No = 82%	Yes = 18% No = 82%	Yes = 21% No = 79%

Table 4. Subject Demographic Information by Round.

conducted as a rapid way to identify the questions with the greatest central tendencies for their constructs. For each of the 13 constructs, the three questions loading highest on their factors were chosen.

Although eight of the 13 constructs loaded on separate factors, as had been hoped, five constructs (information quality, functional fit-to-task, tailored information, innovativeness, and relative advantage) did not, at least in the EFA. This might have been indicative of discriminant validity problems, but that was not an important concern at this point for two reasons. In the first place, EFA is unstable, especially when it contains many marginal questions. As Bagozzi and Phillips have shown, EFA has serious deficits in assessing discriminant validity [9]. Not only is there no explicit statistical test that can be used, but the factor loadings are unpredictable and can be quite misleading. Bagozzi and Phillips give an example where an EFA collapsed factors that were shown to be distinct in a Confirmatory Factor Analysis (CFA), similar to the present results [9, pp. 477–479]. Second, as previously planned, a CFA analysis (a much more powerful test of discriminant validity than EFA) would soon be used on the remaining questions.

Discriminant Validity: Round 2. The overall fit of the model was tested by confirmatory factor analysis (using AMOS). For an overall test of discriminant validity, following Anderson and Gerbing and also Bagozzi and Phillips, the discriminant validity of each of the 78 possible pairs of the 13 constructs was tested individually [5, 9]. For each pair, the fit of the full model was compared with the fit of a model that was exactly the same except for collapsing the pair of constructs into one construct. More specifically, for the collapsed model, the two target constructs were constrained to have a correlation of 1.0 and equal correlations with the 11 other constructs. The general idea is illustrated in Figure 2. Since the full model and the collapsed model are "hierarchically
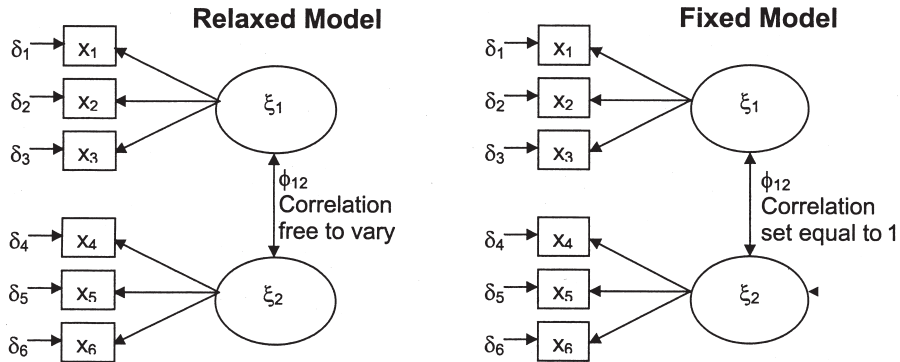
Figure 2. Fixed vs. Relaxed Model Comparison for Assessing Discriminant Validity

nested," a chi-square difference test makes it possible to see whether relaxing the restriction results in a statistically significant improvement in the fit [89].

With so many comparisons (78 possible pairs) and a significance level of 0.05, it would be possible for random variations to influence the results at least in some cases. Therefore, in interpreting the results, consistent with Anderson and Gerbing, it was resolved that "respecification decisions should not be based solely on statistical considerations alone, but rather in conjunction with theory and content considerations. . . . [This] reduces the possibility of taking advantage of sampling error to attain goodness of fit" [5, p. 416]. In addition, the data sets from rounds 3 and 4 could be used to confirm assessments of discriminant validity.

Every possible pair of constructs was found to be distinct at the 0.05 level of significance, with one exception: response time versus intuitive operations. This pair missed the 0.05 significance cutoff ($\chi^2 = 19.95$, $df = 12$, $p < 0.07$). Because these two constructs were so different conceptually, it seemed unwise to conclude that they were a single concept, and both were retained for further testing. As will be seen, the analysis of the data from rounds 3 and 4 vindicated this faith in the discriminant validity of these two constructs.

Statistical significance in discriminant validity tests does not necessarily imply that the distinction between constructs is practically significant [5]. Vandenberg and Lance address this concern by proposing that in addition to looking at chi-square differences between the nested models as the definitive test, researchers might also look at changes in the comparative fit index (CFI) [100]. They suggest that an improvement of 0.02 or more in the CFI is a definite indication of the practical significance of discriminant validity.

This approach was not used because of a feeling that Vandenberg and Lance's approach was biased toward smaller models, where a local improvement was more likely to have a noticeable affect on overall fit. Therefore, for the tests of practical significance, the four pairs of constructs with the highest correlations were looked at, and each pair was tested in isolation (i.e., two constructs and six questions). The testing followed Vandenberg and Lance's suggestion that CFI should increase by at least 0.02 for the discriminant valid-

ity to be "practically significant" [100]. In these tests, three of the four pairs achieved an improvement in CFI of 0.02 or greater (information quality and tailored communication, on-line completeness and relative advantage, visual appeal and innovativeness), and one pair (information quality and functional fit-to-task) did not (change in CFI of 0.01). Recall that there were similar concerns about the discriminant validity of these two constructs based on the analysis of the round 1 data. Reexamining the questions and the definitions of these two constructs led to the conclusion that in fact there was quite a bit of overlap and little practical significance in maintaining them as separate constructs. Therefore, they were collapsed into a single construct.

All together, the round 2 analyses reduced WebQual from 13 to 12 constructs, and from 83 to 36 questions. See Appendix 2 for the final items by construct.

Stage 4: Confirmatory Assessment of Validity

A complete, confirmatory assessment of the measurement validity of the final version of the questionnaire was performed with two new samples, round 3 containing 311 students collected in the summer/fall 2000 time frame and round 4 containing 377 students collected in the 2002/2003 time frame.[5] With these two new samples, all the empirically testable concerns suggested by Bagozzi in Table 1 were addressed (the last four rows).

Table 3 displays the overall fit of the CFA measurement model (using AMOS), for round 2, round 3, and round 4. The four recommended fit indices for all the rounds indicate quite acceptable fit for the final version [59, 100]. The standardized item loadings on their constructs ranged from 0.97 to 0.54. However, all but two in round 3 and all but one in round 4 were 0.64 or above. These results indicate strong support for the overall fit of the model. Correlations for the constructs in rounds 3 and 4 are found in Appendix 3.

Internal Consistency (Reliability)

The reliability of the final questionnaire (12 constructs with three questions each) was calculated using Cronbach's alpha [21]. As Table 5 shows, for round 3, the alphas of the 12 constructs ranged from 0.72 to 0.93, with 10 of the 12 constructs having an alpha greater than 0.80. Of the two remaining constructs, one had an alpha of 0.79, and one had an alpha of 0.72. The results were almost identical for round 4. Since the suggested hurdle value for frequently used instruments ranges from alpha levels of 0.60 to 0.80 [71], all were above the more generous hurdle of 0.60, and only two constructs (on-line completeness and tailored communications) were slightly below the more restrictive hurdle.

Discriminant Validity

Discriminant validity was checked separately for round 3 and round 4. All possible pairs of the remaining constructs (66 possible pairs from 12 constructs

Construct	2000 summer/fall	2002/2003
Informational fit-to-task	0.86	0.85
Tailored information	0.80	0.78
Trust	0.90	0.93
Response time	0.88	0.80
Ease of understanding	0.83	0.81
Intuitive operations	0.79	0.83
Visual appeal	0.93	0.91
Innovativeness	0.87	0.87
Emotional appeal	0.81	#
Consistent image	0.87	0.83
On-line completeness	0.72	0.71
Relative advantage	0.81	0.80

Table 5. Construct Reliabilities for Rounds 3 and 4—Cronbach's Alpha.

Note: Emotional appeal was not included; a different version of the emotion questions was used for the 2002/2003 data collection.

in round 3 and 55 possible pairs from 11 constructs in round 4)[6] were tested to see if fit was improved when any pair was collapsed into a single construct, using the same approach as described for round 2. Again it was recognized that with so many tests (121 in the two rounds) some would fail to achieve significance through random chance.

Recall that the pair comprising response time and intuitive operations had not quite achieved discriminant validity in round 2. In the round 3 data, these two constructs were shown to be distinct at the 0.0001 level of significance.

All other pairs of constructs in round 3 were distinct at least at the 0.05 level of significance, with two exceptions: innovativeness versus emotional appeal, and information fit-to-task versus tailored information. In tests of the round 4 data, all pairs of constructs were distinct at the 0.05 level of significance, with one exception—visual appeal and innovativeness–which missed significance (0.059). To summarize the round 2, round 3, and round 4 discriminant validity results, out of 187 pair tests across three rounds of data, 184 showed statistically significant discriminant validity, and four missed achieving significance. Even for those four pairs of constructs, each pair showed statistically significant discriminant validity in at least two of the three rounds of tests.

Finally, the data from rounds 2, 3, and 4 were put together, repeating the chi square difference tests for discriminant validity. All 55 pairs of constructs were distinct ($p < 0.02$). All in all, it can be confidently stated that the 12 constructs are separate characteristics of a Web site in the minds of consumers.

Convergent Validity

Strictly speaking, convergent validity (à la Bagozzi) refers to the correlation between the new measure and a second "maximally different" measure of the same construct [7, 8].[7] For example, a measure of performance using managerial perceptions might be compared to an objective measure of performance or

to some other already accepted measure of performance as a way to validate the new measure. Since there were 12 constructs, it would be necessary to find alternative measures for all 12 constructs. This has not been done, to date.

More general measures of convergent validity are possible. Following the example of the development of SERVQUAL [73], subjects were asked one additional question evaluating the overall Web site—"Overall, how would you rate the quality of this Web site?" (1- to 7-point scale anchored on "poor" and "excellent"). The word "quality" is associated with such words as "goodness," "excellence," "value," and "worth" in *Roget's Thesaurus* (§674.1), and thus this single question captures the central essence of a consumer's overall evaluation of a Web site.

The total WebQual score (36 items) was computed and then compared to the overall Web site question. The two were correlated at 0.78 ($p < 0.001$), giving an overall and very general indication of convergent validity.

Nomological Validity/Predictive Validity

Nomological validity is the ability of a new measure to perform as expected in a network of known causal relations and well-established measures [7, 86, 106]. One cannot have confidence in a measure if it does not behave in a reasonable fashion in relation to other accepted constructs. In the case of WebQual, a relevant demonstration of nomological validity would be the extent to which constructs that have been conceptually linked to Web site use (in this and other research papers) do actually predict a Web visitor's behavioral intention in ways that are consistent with current knowledge and expectations.

One way to test WebQual's ability to predict consumer intentions would be to regress the 12 constructs on intent to reuse/purchase. However, since some of the 12 constructs are highly correlated, multicollinearity would be a problem with a regression analysis of this kind.

The multicollinearity problem can be side-stepped by a second-order factor analysis in advance of looking at the structural links between the second-order factors and intent to reuse. Because this is a significant piece of research in its own right, it is beyond the scope of the current study. However, one can take advantage of such an analysis to give the groupings and loadings from the 12 underlying constructs to five second-order constructs [63]. The 12 underlying and five second-order constructs are shown at the left-hand side and middle of Figure 3, with the loadings being the single numbers next to arrows. As can be seen, the 12 first-order factors resolve into five second-order factors: usefulness, trust, response time, ease of use, and entertainment.[8] These groupings and loadings were used to analyze the links to intention for three separate datasets: round 2 (spring 2000, $N = 336$), round 3 (summer/fall 2000, $N = 311$), and round 4 (2003/2003 $N = 377$). All fit indices were found to be acceptable (*see Table 3*).

Intention to reuse/purchase was measured using the questions shown in Appendix 2. Note the expansion from the use of two indicators in rounds 2 and 3 (Cronbach's alpha of 0.90 and 0.84, respectively) to five indicators in round 4 (Cronbach's alpha of 0.94). The two-indicator measure and the five-indicator measure are both quite adequate. The different measures were not
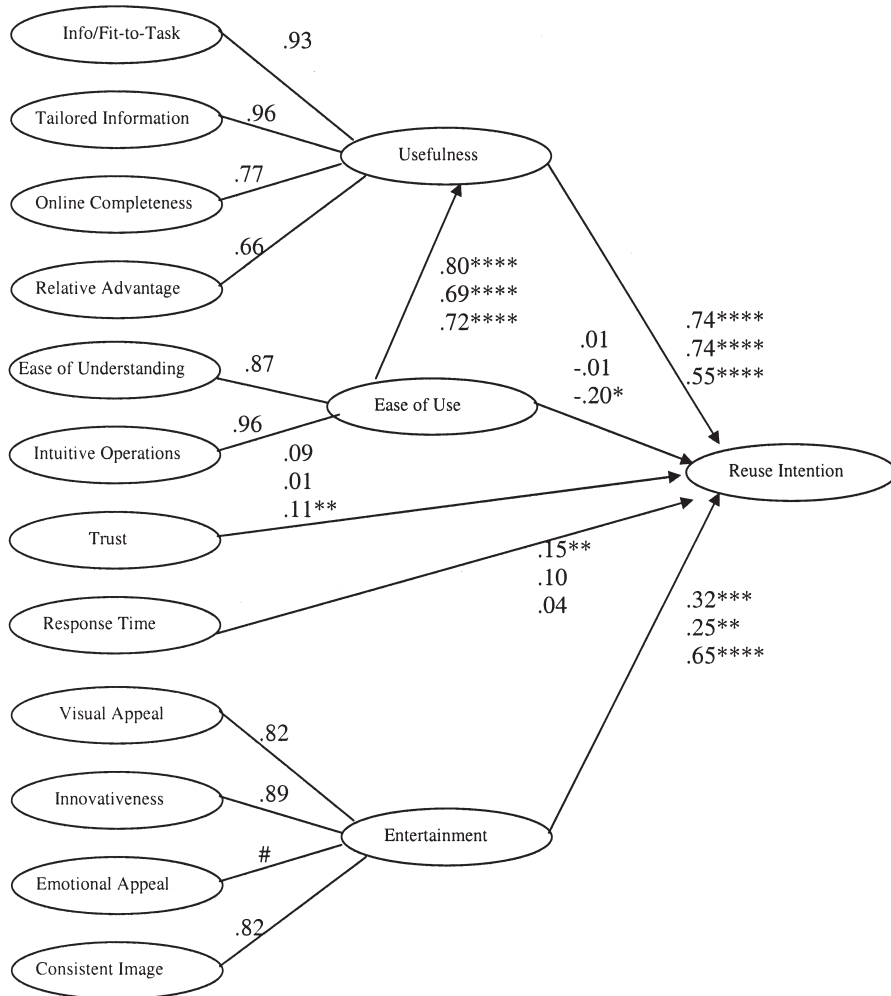
Figure 3. Nomological Validity: Test of Links Between Five Second-Order Constructs and Reuse

Notes: Arrows leading into Reuse and Usefulness show three values, the first is from 2000 Spring, second is from 2000 Summer/Fall, and the third is from 2002/2003. # Emotional Appeal was removed from this analysis in order to make the three data sets consistent—a different version of the emotion questions were used for the 2002/2003 data collection. **** = significance at 0.001; *** = significance at 0.005; ** = significance at 0.010; * = significance at 0.050.

combined into a single analysis, but were considered to be roughly equivalent for the nomological validity analysis.

Figure 3 and Table 6 show the results of the estimated causal links, by displaying three numbers for each causal path, corresponding to the spring 2000, the summer/fall 2000, and the 2002/2003 time periods. Having three separate samples spread across three time periods makes it possible to see some interesting results. In all three time periods, usefulness and entertainment have strongly significant links to intent to reuse. Further, in the spring 2000 time

Date of data collection	Round 2 2000 spring	Round 3 2000 summer/fall	Round 4 2002/ 2003
Predicting reuse			
N	335	311	377
Adj. R^2	0.56****	0.35****	0.68****
Usefulness	0.74****	0.74****	0.55****
Entertainment	0.32***	0.25**	0.65****
Trust	0.09	0.01	0.11**
Response time	0.15**	0.10	0.04
Ease of use	0.01	-0.01	-0.20*
Predicting usefulness			
Adj. R^2	0.45****	0.47****	0.59****
Ease of use	0.80****	0.69****	0.72****
Average subject characteristics			
Years on Web	4.7	4.8	6.2
Ever made a Web purchase	Yes (66%)	Yes (64%)	Yes (73%)
Web purchases/last month	1.71	0.81	1.12
Heard of company	Yes (18%)	Yes (18%)	Yes (21%)

Table 6. Nomological Validity: Predicting Reuse and Usefulness from Second-Order Factors (Using Two Confirmatory Datasets Round 3 from 2000 Summer/Fall and Round 4 from 2002/2003, plus Round 2 Spring 2000 Data for Comparison).

**** = significance at 0.001; *** = significance at 0.005; ** = significance at 0.010; * = significance at 0.050.

period, response time is linked to reuse but trust is not, while in the 2002/2003 time period, trust is linked to reuse at the 0.01 level of significance, but response time is not.

The importance of usefulness is consistent with findings by Davis and others. The importance of entertainment is also consistent with previous research [101]. The appearance of trust and response time as dimensions reveals, as was suggested at the beginning of this paper, that Web use has predictors beyond the usefulness and ease of use suggested by TAM. The results suggest the intriguing possibility that the relative importance of entertainment may be increasing over time. If these shifts are confirmed by future research, one explanation might be that the response time in the earlier data was a factor of less mature technologies and over several years became much less of a factor.

Although trust has a significant link to reuse in round 4, the lack of that link in rounds 2 and 3 is somewhat surprising. One possible explanation is that with more experience and a higher rate of Web purchases in the later time period, these consumers were more likely to have encountered Web sites where trust is an issue, and thus were more sensitized to it. Perhaps consumers are more aware today than they were several years ago of the danger of purchasing at untrustworthy Web sites, but since this was not hypothesized, it is merely speculation at this point.

The lack of a positive linkage between ease of use and intent to reuse is not totally unexpected. It is consistent with the findings of Davis and other that for experienced users, ease of use is not a direct predictor of intent to reuse

but is a predictor of usefulness [26, 97, 105]. As shown at the bottom of Table 6, the subjects were quite experienced—they had many years of experience on the Web. Although the significant negative link between ease of use and reuse in round 4 is surprising, it is consistent with a multicollinearity relationship between ease of use and usefulness—these two constructs are highly correlated with each other (0.77), and both are quite correlated with reuse.[9] In this case, the negative link in round 4 should not be viewed as a causal link. Because ease of use is still an important construct, and consistent with the findings of Davis and his colleagues, the link between ease of use and usefulness has been included in the present analysis and is shown in Figure 3 and the middle portion of Table 6. Overall, these results are strong positive confirmation of nomological validity.

Discussion

Implications for Researchers

The primary research contribution is the development of a validated instrument for evaluating Web sites by following a more rigorous process than predecessors and revealing a wider range of dimensions. Measurement instruments are in a very real sense a research "infrastructure." The development of the measure in itself may do little to prove or disprove a specific hypothesized relationship, and so may not move theory forward very much. However, when an important domain figures prominently in a variety of research programs, the development of a more effective measurement instrument moves many research programs forward. As noted earlier in this paper, five articles in a recent special issue of *ISR* developed instruments to measure user reactions to Web sites, and none of these measures met the suggested requirements for a general purpose instrument to capture consumer evaluations of a Web site. When there is no carefully developed and tested general-purpose instrument available, each researcher is forced to spend research time and effort on developing a measure, with the result that measures are weaker, and researcher time is drawn away from the theoretical interest of the research.

The importance of instrument development in research is easily seen by looking at the citation counts of three papers introducing two prominent measurement instruments: User Information Satisfaction and SERVQUAL [10, 53, 73]. These three papers focused almost exclusively on the development and testing of the two instruments, not on theory testing. According to the Web of Science, the paper by Bailey and Pearson has been cited 341 times, the one by Ives, Olson, and Baroudi has been cited 371 times, and the one by Parasuraman, Zeithaml, and Berry has been cited 601 times. These instruments have had a huge impact on the research community.

A second contribution is that, by starting with TAM and looking both within and beyond the original constructs of usefulness and ease of use, this paper demonstrates how to apply that theory base to additional technology contexts. This approach may be valuable as new information technologies appear on the scene (Web-enabled handheld devices, navigation systems, etc.) As this

evolution occurs, so will the need for IS and marketing research to meld in order to measure the effectiveness of the new devices. WebQual stands as a significant contribution to the current marketing and IS literature.

A third research contribution is WebQual's comprehensive scope, which takes in the many different factors that contribute to the reuse of Web sites. This facilitates at least a first attempt at discerning their relative importance.

Implications for Practice

A poor Web site can cause bad press, customer dissatisfaction, and even customer loss [46]. WebQual offers practitioners a validated reliable instrument for evaluating B2C Web sites. The 36 items provide a valuable and accurate measure of 12 constructs that affect consumer intention to purchase and revisit, a critical pragmatic indicator of a site's value. Companies can use WebQual to "consumer test" sites under development and detect which elements need improvement prior to public release. Given the investment many companies make in their Web sites, a cost-effective instrument for comparing design alternatives should have substantial practical value.

Both purely Web-based and click-and-mortar companies can gain from conducting a WebQual analysis. For example, a click-and-mortar firm such as Barnes and Noble can benefit from understanding how effective each aspect of its Web site is, and so too does Amazon, a purely Web-based competitor. The two companies have different Web strategies, but both gain from knowing how well they rate on the basic Web site evaluation constructs, such as intuitive operations and informational fit to task. In addition, WebQual also serves as a means of benchmarking against competitors. Barnes and Noble may determine, given its business strategy, that it is willing to score lower on certain components of WebQual, such as relative advantage, but not on on-line completeness or consistent image.

One of the values of WebQual is that it provides a fine-grained analysis of a site's shortcomings. While an instrument based on TAM might be able to diagnose a shortfall on ease of use, WebQual will indicate whether the ease of use problem arises from problems with ease of understanding or with intuitive operations. Knowing that a Web site is lacking on a particular dimension would give the company a clearer focus for corrective action.

Limitations and Future Research

WebQual's development was based on the responses of undergraduate business students to a selected group of Web sites. These subjects were typical of a substantial body of Web consumers but were not necessarily representative of all consumers. Furthermore, many of the subjects were not ongoing customers of the sites selected for assessment. However, firms are interested both in servicing existing customers and in attracting new customers, so they need an instrument that is valid for both groups. These important limitations

are typical of those facing most instrument developers because such work often needs to start in an environment where many subjects are readily and repeatedly available. Future research can explore the implications for Web site quality with after-sales service quality added. Further confirmatory research needs to be done with broad samples of consumers for a range of Web sites. Also, while WebQual indicates where a site may have deficiencies, follow-up research is needed that focuses on the link between the specific changes that can be made to a Web site and the resulting changes in customer perceptions. Given that WebQual has completed initial development and refinement, it is now appropriate to investigate some other important issues, such as business-to-business (B2B) and noncommercial-to-consumer (N2C) sites [31]. In addition, new technologies are ever expanding and broadening the capabilities of Web sites. Subsequent WebQual research will need to fold in additional elements, such as audio, video, and mobile Web applications [37]. Other tasks beyond information gathering, as used in this study, may reveal a different breakdown of the constructs. For example, information quality and functional fit-to-task may not combine as they did in this research.

Finally, the study looks at traditional uses of the Web to convey information and facilitate transactions. However, other ways to utilize the Web are evolving, such as networking and blogging. These types of activities are not the focus of the current WebQual research but may be included in future research.

Conclusion

This study shows how the TRA and TAM can be applied to a new domain, thereby enabling us to see the relative importance of many different characteristics in predicting reuse of a Web site. Its major contribution, however, is in providing useful "research infrastructure" for the domain of the electronic market space. In the age of the Internet and electronic commerce, IS and marketing researchers and practitioners need a means of assessing Web sites. The research effort described in this paper has produced a valid and reliable instrument for consumer evaluations of Web sites. WebQual can be freely used by academics without permission. It should be able to support a wide range of important IS and marketing studies as researchers attempt to understand a variety of issues related to the electronic market space. Although individual researchers may focus on somewhat different issues, and may need to develop one or more supplemental measures for their own specialized constructs, the existence of a validated baseline measure should be a major advantage. It provides a starting point that will speed and enhance the quality of measurement for many studies. In addition, to the extent that researchers converge on standard measures, it will be much easier to compare and consolidate findings across studies.

Of course, WebQual is not the final word in measurement instruments for Web sites. Like any measurement instrument for an important research domain, it will certainly be modified, added to, and refined in various ways.

NOTES

1. Because of the instrument's similar name, it is referred to as WQ.

2. The specific conditions not met are these: 2 [72], 3 [58], 2 and 3 [3], 6 [29], 7 [61], and phase 1 of 6, phase 3 of 3 [65]. More detail is available in Appendix 1.

3. A selection of MIS graduate and undergraduate students, each of whom had at least some experience on the Web.

4. All the items in all the questionnaires in all the research domains were originally generated by the researchers. For this study, items generated after round 1 were tested and validated by three additional data-collection rounds.

5. One company raised concerns about the questions for emotional appeal ("I feel happy . . .," "I feel cheerful . . .," etc.), so a different approach was adopted in round 3. The new questions did not perform well, so the old questions were restored. However, the emotional measures are missing from the analyses for round 4.

6. The emotional appeal questions were excluded from the analysis for round 4 because their wording had changed for that round. However, this construct was found to be significantly different from the other 11 constructs in rounds 2 and 3.

7. Note that this is distinct from reliability, which assesses whether all items of a particular construct correlate strongly with other "maximally similar" items, and by assumption, with the underlying construct.

8. AMOS was used to assess the structural model, and all first-order constructs were allowed to be correlated.

9. According to a personal conversation (1990) between one author and John Neter, multicollinearity is strongly indicated when one variable is a strong regression predictor of another, but adding an additional independent variable to the regression causes the sign of the relationship to change. This was exactly the case with ease of use. A regression without usefulness shows a strong link (0.19, significant at 0.01) between ease of use and reuse. Adding usefulness to the regression causes the link between ease of use and reuse to completely reverse its direction (–0.20, significant at 0.05). This is strong evidence that the apparent negative link is due to multicollinearity and is not a true causal link.

REFERENCES

1. Aaker, D.A., and Stayman, D.M. Measuring audience perceptions of commercials and relating them to ad impact. *Journal of Advertising Research, 30,* 4 (1990), 7–17.

2. Agarwal, R., and Karahanna, E. Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *Management of Information Systems Quarterly, 24,* 4 (2000), 665–694.

3. Agarwal, R., and Venkatesh, V. Measuring factors that influence the success of Internet commerce. *Information Systems Research, 13,* 2 (2002), 168–186.

4. Ajzen, I., and Fishbein, M. *Understanding Attitudes and Predicting Social Behavior.* Englewood Cliffs, NJ: Prentice Hall, 1980.

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

6. Babbie, E. *The Practice of Social Research,* 8th ed. Belmont, CA: Wadsworth, 1998.

7. Bagozzi, R.P. *Causal Models in Marketing.* New York: John Wiley, 1980.

8. Bagozzi, R.P., and Edwards, J.R. A general approach for representing constructs in organizational research. *Organizational Research Methods, 1,* 1 (1988), 45–87.

9. Bagozzi, R.P., and Phillips, L.W. Representing and testing organizational theories: A holistic construal. *Administrative Science Quarterly, 27,* 3 (1982), 459–489.

10. Bailey, J.E., and Pearson, S.W. Development of a tool measuring and analyzing computer user satisfaction. *Management Science, 29,* 5 (1983), 530–544.

11. Barnes, S.J., and Vidgen, R. An evaluation of cyber-bookshops: The WebQual method. *International Journal of Electronic Commerce, 6,* 1 (2001), 11–30.

12. Baroudi, J.J., and Orlikowski, W.J. A short-form measure of user information satisfaction: A psychometric evaluation and notes on use. *Journal of Management Information Systems, 4,* 4 (spring 1988), 44–59.

13. Benbunan-Fich, R. Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management, 39,* 2 (2001), 151–163.

14. Berthon, P.R.; Pitt, L.F.; and Watson, R.T. Re-surfing W3: Research perspectives on marketing communication and buyer behaviour on the World Wide Web. *International Journal of Advertising, 15,* 4 (1996), 287–301.

15. Bossert, J.L. *Quality Function Deployment: A Practitioner's Approach.* Milwaukee: ASQC Quality Press, 1991.

16. Burns, E. The online battle of the sexes (www.clickz.com/showPage.html?page=3574176/).

17. Campbell, D.T., and Fisk, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 2 (1959), 81–105.

18. Cappiello, C.; Francalanci, C.; and Pernici, B. Time-related factors of data quality in multichannel information systems. *Journal of Management Information Systems, 20,* 3 (winter 2003–4), 71–91.

19. Cheung, W.; Chang, M.K.; and Lai, V.S. Prediction of Internet and World Wide Web usage at work: A test of an extended Triandis model. *Decision Support Systems, 30,* 1 (2000), 83–100.

20. Cook, T.D., and Campbell, D.T. *Quasi-Experimental Design: Design and Analysis Issues for Field Settings.* Skokie, IL: Rand-McNally, 1979.

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

22. Cronbach, L.J., and Meehl, P.C. Construct validity in psychological tests. *Psychological Bulletin, 52,* 4 (1955), 281–302.

23. Csikszentmihalyi, M. *Beyond Boredom and Anxiety,* 2d ed. San Francisco: Jossey-Bass, 1977.

24. Cyberatlas. Reasons visitors return to sites (www.clickz.com).

25. Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly, 13,* 3 (1989), 319–339.

26. 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), 982–1003.

27. DeLone, W.H., and McLean, E.R. Information systems success: The quest for the dependent variable. *Information Systems Research, 3,* 1 (1992), 60–95.

28. DePelsmacker, P., and VanDenBergh, J. The communication effects of provocation in print advertising. *International Journal of Advertising, 15,* 3 (1996), 203–221.

29. Devaraj, S.; Ming, F.; and Rajiv, H. Antecedents of B2C channel satisfaction and preference: Validating e-commerce metrics. *Information Systems Research, 13,* 2 (2002), 316–333.

30. Doney, P.M., and Cannon, J.P. An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing, 61,* 2 (1997), 35–51.

31. Driedonks, C.; Gregor, S.; Wassenaar, A.; and Van Heck, E. Economic and social analysis of the adoption of B2B electronic marketplaces: A case study in the Australian beef industry. *International Journal of Electronic Commerce, 9,* 3 (2005), 49–72.

32. Ducoffe, R. How consumers assess the value of advertising. *Journal of Current Issues and Research in Advertising, 17,* 1 (1995), 1–18.

33. Eighmey, J. Profiling user responses to commercial Web sites. *Journal of Advertising Research, 37,* 3 (1997), 59–66.

34. Elliot, M., and Speck, P. Consumer perceptions of advertising clutter and its impact across various media. *Journal of Advertising Research, 38,* 1 (1998), 29–41.

35. Fishbein, M., and Ajzen, I. *Beliefs, Attitude, Intention, and Behavior: An Introduction to Theory and Research.* Reading, MA: Addison-Wesley, 1975.

36. Galletta, D.F.; Henry, R.; McCoy, S.; and Polak, P. Sensitivity to Web delays: A contingency analysis of users. *Journal of the Association of Information Systems, 5,* 1 (2004), 1–28.

37. Gebauer, J., and Shaw, M. Success factors and impacts of mobile business applications: Results from a mobile e-procurement study. *International Journal of Electronic Commerce, 8,* 3 (2004), 19–41.

38. Gefen, D.; Karahanna, E.; and Straub, D.W. Trust and TAM in online shopping: An integrated model. *Management Information Systems Quarterly, 27,* 1 (2003), 51–90.

39. Geissler, G.L.; Zinkhan, G.M.; and Watson, R.T. Web home page complexity and communication effectiveness. *Journal of the Association of Information Systems, 2,* 2 (2001), 288–302.

40. George, J.M. State of trait: Effects of positive mood on prosocial behaviors at work. *Journal of Applied Psychology, 76,* 2 (1991), 299–307.

41. Ghose, S., and Dou, W. Interactive functions and their impacts on the appeal of Internet presence sites. *Journal of Advertising Research, 38,* 2 (1998), 29–43.

42. Goodhue, D. Development and measurement validity of a task-technology fit instruments for user evaluations of information systems. *Decision Sciences, 29,* 1 (1998), 105–138.

43. Goodhue, D.L., and Thompson, R.L. Task-technology fit and individual-performance. *Management of Information Systems Quarterly, 19,* 2 (1995), 213–236.

44. Gournaris, S., and Sergios, D. Assessing service quality on the Web: Evidence from business-to-consumer portals. *Journal of Service Marketing, 17,* 4/5 (2003), 526.

45. Griffith, D.A.; Krampf, R.E.; and Palmer, J.W. The role of interface in electronic commerce: Consumer involvement with print versus on-line catalogs. *International Journal of Electronic Commerce, 5,* 4 (summer 2001), 135–153.

46. Gruman, G. E-commerce blurs lines of integrity, but they still exist. *Computerworld, 33,* 8 (1999), 37.

47. Ha, L., and Litman, B.R. Does advertising clutter have diminishing and negative returns? *Journal of Advertising, 26,* 1 (1997), 31–42.

48. Hair, J.F., Jr.; Anderson, R.E.; Tatham, R.L.; and Black, W.C. *Multivariate Data Analysis,* 5th ed. Upper Saddle River, NJ: Prentice Hall, 1998.

49. Hensley, R.L. A review of operations management studies using scale development techniques. *Journal of Operations Management, 17,* 3 (1999), 343–358.

50. Hinkin, T.R. A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1,* 1 (1998), 104–121.

51. Hoffman, D.L.; Novak, T.P.; and Peralta, M. Building consumers trust online. *Communications of the Association for Computer Machinery, 42,* 4 (1999), 80–85.

52. Hong, W.; Thong, J.Y.L.; Mong, W.-M.; and Tam, K.-Y. Determinants of user acceptance of digital libraries: An empirical examination of individual differences and system characteristics. *Journal of Management Information Systems, 18,* 3 (2001–2), 97–124.

53. Ives, B.; Olson, M.H.; and Baroudi, J.J. The measurement of user information satisfaction. *Communications of the Association for Computer Machinery, 26,* 10 (1983), 785–793.

54. Johnson, C. *US eCommerce: 2005 to 2010: A Five-Year Forecast and Analysis of US Online Retail Sales,* Forrester Research, September 14, 2005.

55. Katerattanakul, P., and Siau, K. Measuring information quality of Web sites: Development of an instrument. In *Proceedings of the 20th International Conference on Information Systems.* Atlanta: AIS, 1999, pp. 279–285.

56. Kaynama, S., and Black, C.I. A proposal to assess the service quality of online travel agencies: An exploratory study. *Journal of Professional Services Marketing, 21,* 1 (2000), 63–88.

57. Kettinger, W.J., and Lee, C.C. Pragmatic perspectives on the measurement of information systems service quality. *Management of Information Systems Quarterly, 21,* 2 (1997), 223–240.

58. Kim, J.; Lee, J.; Han, K.; and Lee, M. Business as buildings: Metrics for the architectural quality of Internet business. *Information Systems Research, 13,* 2 (2002), 239–254.

59. Kline, R.B. *Principles and Practices of Structural Equation Modeling.* New York: Guilford Press, 1998.

60. Kotler, P. Atmospherics as a marketing tool. *Journal of Retailing, 49,* 4 (1973), 48–64.

61. Koufaris, M. Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research, 13,* 2 (2002), 205–223.

62. Lee, Y.; Strong, D.M.; Kahn, B.K.; and Wang, R.Y. AIMQ: A methodology for information quality assessment. *Information & Management, 40,* 2 (2002), 133–146.

63. Loiacono, E.; Chen, D.; and Goodhue, D. WebQual™ revisited: Predicting the intent to reuse a Web site. Paper presented at the Americas Conference on Information Systems (AMCIS), Dallas, Texas, August 9–11, 2002.

64. Machlis, S. Low-tech marketing. *Computerworld, 33,* 3 (1999), 41.

65. McKinney, V.; Yoon, K.; and Zahedi, F.M. The measurement of Web customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research, 13,* 2 (2002), 296–315.

66. McKnight, D.H., and Chervany, N.L. What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce, 6,* 2 (winter 2001–2), 35–59.

67. McKnight, D.H.; Choudhury, V.; and Kacmar, C. Developing and validating trust measures for e-commerce: An interactive typology. *Information Systems Research, 13,* 3 (2002), 334–361.

68. Moore, G., and Benbasat, I. Development of an instrument to measure perceptions of adopting an information technology innovation. *Information Systems Research, 2,* 3 (1991), 192–222.

69. Nielsen, J. Change the color of visited links (www.useit.com/alertbox/20040503.html).

70. Novak, T.P.; Hoffman, D.L.; and Yung, Y. Modeling the structure of the flow experience among Web users: A structural modeling approach. Paper presented at the Association for Consumer Research Conference, Columbus, OH, September 30–October 3, 1999.

71. Nunnally, J.C. *Psycholometric Theory.* New York: McGraw-Hill, 1978.

72. Palmer, J.W. Web site usability, design, and performance metrics. *Information Systems Research, 13,* 2 (2002), 151–167.

73. Parasuraman, A.; Zeithaml, V.A.; and Berry, L.L. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64,* 1 (1988), 12–40.

74. Pavlou, P.A. Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce, 7,* 3 (2003), 101–134.

75. Pennington, R.; Dixon Wilcox, H.; and Grover, V. The role of system trust in business-to-consumer transactions. *Journal of Management Information Systems, 20,* 3 (winter 2003–4), 197–226.

76. Pine, B.J., III, and Gilmore, J.H. Welcome to the experience economy. *Harvard Business Review, 76,* 4 (1998), 97–105.

77. Ramaswami, S.N.; Strader, T.J.; and Brett, K. Determinants of on-line channel use for purchasing financial products. *International Journal of Electronic Commerce, 5,* 2 (winter 2000–2001), 95–118.

78. Raney, A.A.; Arpan, L.M.; Pashupati, K.; and Brill, D.A. At the movies, on the Web: An investigation of the effects of entertaining and interactive Web content on site and brand evaluation. *Journal of Interactive Marketing, 17,* 4 (2003), 38–53.

79. Ratchford, B.T.; Talukdar, D.; and Lee, M. A model of consumer choice of the Internet as an information source. *International Journal of Electronic Commerce, 5,* 3 (spring 2001), 7–21.

80. Resnik, A., and Stern, B.L. An analysis of information content in television advertising. *Journal of Marketing, 41,* 1 (January 1977), 50–53.

81. Richins, M. Measuring emotions in the consumption experience. *Journal of Consumer Research, 24,* 2 (1997), 127–146.

82. Rogers, E.M. *Diffusion of Innovations,* 3d ed. New York: Free Press, 1983.

83. Rust, R.T., and Lemon, K.N. E-service and the consumer. *International Journal of Electronic Commerce, 5,* 3 (spring 2001), 85–101.

84. Schubert, P. Extended Web assessment method (EWAM): Evaluation of e-commerce applications from the customer's viewpoint. *International Journal of Electronic Commerce, 7,* 2 (winter 2002–3), 51–80.

85. Schubert, P., and Selz, D. Web assessment: A model for the evaluation and the assessment of successful electronic commerce applications. *International Journal of Electronic Markets, 7,* 3 (1997), 46–48.

86. Sethi, V., and King, W.R. Construct measurement in information systems research: An illustration in strategic systems. *Decision Sciences, 22,* 3 (1991), 455–472.

87. Seybold, P.B., and Marshak, R.T. *Customers.com: How to Create a Profitable Business Strategy for the Internet and Beyond.* New York: Times Business, 1998.

88. Shneiderman, B. CUU: Bridging the digital divide with universal usability. *Interactions, 8,* 1 (March 2001), 11–15.

89. Steiger, J.H.; Shapiro, A.; and Brown, M.W. On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50,* 3 (1985), 253–264.

90. Steinfield, C.; Bouwman, H.; and Adelaar, T. The dynamics of click-and-mortar electronic commerce: Opportunities and management strategies. *International Journal of Electronic Commerce, 7,* 1 (2002), 93–119.

91. Steuer, J. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42,* 4 (1992), 73–93.

92. Straub, D.W. Validating instruments in MIS research. *Management Information Systems Quarterly, 13,* 2 (1989), 147–169.

93. Strong, D.M. IT process designs for improving information quality and reducing exception handling: A simulation experiment. *Information & Management, 31,* 5 (1997), 251–263.

94. Su, L.T.; Chen, H.; and Dong, X. Evaluation of Web-based search engines from the end-user's perspective: A pilot study. Paper presented at the 61st Annual Meeting of the American Society for Information Science, Pittsburgh, PA, October 25–29, 1998.

95. Suh, B., and Han, I. The impact of customer trust and perception of security control on the acceptance of electronic commerce. *International Journal of Electronic Commerce, 7,* 3 (spring 2003), 135–161.

96. Swanson, N.E. Power: A critical systems-development factor. *Information & Management, 9,* 4 (1985), 209–213.

97. Szajna, B. Empirical evaluation of the revised technology acceptance model. *Management Science, 42,* 1 (1996), 85–92.

98. Tan, Y., and Thoen, W. Toward a generic model of trust for electronic commerce. *International Journal of Electronic Commerce, 5,* 2 (winter 2000–2001), 61–74.

99. Todd, P., and Benbasat, I. The use of information in decision-making—An experimental investigation of the impact of computer-based decision aids. *Management of Information Systems Quarterly, 16,* 3 (1992), 373–393.

100. Vandenberg, R., J., and Lance, C.E. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 1 (2000), 4–69.

101. van der Heijden, H. User acceptance of hedonic information systems. *Management Information Systems Quarterly, 28,* 4 (2004), 695–704.

102. Venkatesh, V. Creation of favorable user perceptions: Exploring the role of intrinsic motivation. *Management Information Systems Quarterly, 23,* 2 (1999), 239–260.

103. Venkatesh, V. Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research, 11,* 4 (2000), 342–365.

104. Venkatesh, V., and Speier, C. Computer technology training in the workplace: A longitudinal investigation of the effect of mood. *Organizational Behavior and Human Decision Processes, 79,* 1 (1999), 1–28.

105. Venkatesh, V.; Morris, M.G.; Davis, G.B.; and Davis, F.D. User acceptance of information technology: Toward a unified view. *Management Information Systems Quarterly, 27,* 3 (2003), 425–478.

106. Venkatraman, N., and Grant, J.H. Construct measurement in organizational strategy research: A critique and proposal. *Academy of Management Review, 11,* 1 (1986), 71–87.

107. Wand, Y., and Wang, R.Y. Anchoring data quality dimensions in ontological foundations. *Communications of the Association of Computer Machinery, 39,* 11 (1996), 86–95.

108. Wang, R.Y., and Strong, D.M. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12,* 4 (1996), 5–33.

109. Watson, R.T.; Zinkhan, G.M.; and Pitt, L.F. Integrated Internet marketing. *Communications of the Association of Computer Machinery, 43,* 6 (2000), 97–102.

110. Xie, M.; Wang, H.; and Goh, T.N. Quality dimensions of Internet search engines. *Journal of Information Science, 24,* 5 (1998), 365–372.

111. Yoo, B., and Donthu, N. Developing a scale to measure perceived quality of an Internet shopping site (SITEQUAL). *Quarterly Journal of Electronic Commerce, 2,* 1 (2001), 31–46.

112. Zhuang, Y., and Lederer, A.L. An instrument for measuring the business benefits of e-commerce retailing. *International Journal of Electronic Commerce, 7,* 3 (spring 2003), 65–99.

Appendix 1

Study	WebQual	Palmer 2002 [72]	Koufaris 2002 [61]	Kim et al. 2002 [58]	McKinney et al. 2002 [65]	Devaraj et al. 2002 [29]
Overview	36 Qs on 12 constructs from wide scan of IS and marketing literature	12 Qs on 5 constructs from usability and design; media richness	20 Qs on 5 constructs from TAM (2) and Flow (3)	34 Qs on 6 constructs from architecture (firmness, utility, and delight)	45 Qs on 11 constructs from information quality (6), system quality (5)	32 Qs on 9 constructs from TAM (2), Trans Cost Anal (3), ServQual (4)
Measurement validity concern		The five constructs are factor groupings of questions measuring distinct constructs, not unidimensional.	All 280 subjects rated the same Web site (BooksAMillion)	Did not do required tests of discriminant validity	Ph. 1: respondents did not view any Web sites. Ph. 2, did not do required tests of discriminant validity.	Half the questions were about Web shopping in general, not a specific Web site.
Constructs	Information fit to task	Content richness		Information gathering (variety, accuracy, understandable)	Scope; relevance; adequacy; reliability; usefulness	Uncertainty (could get relevant info)
	Tailored information On-line completeness Relative advantage	Interactivity	Usefulness (improves shopping performance)			Usefulness (of on-line shopping); time (on-line is faster)
	Ease of understanding	Navigation and organization			Understandability	Ease of use of on-line shopping
	Intuitive operations		Perceived control (in control, not confused); ease of use	System interface (easy to learn, delightful, stylish, consistent)	Navigation; usability	
	Visual appeal Innovativeness				Entertainment	

Study	Palmer 2002 [72]	Koufaris 2002 [61]	Kim et al. 2002 [58]	McKinney et al. 2002 [65]	Devaraj et al. 2002 [29]
WebQual					
Emotional appeal		Shopping enjoyment; concentration (how deeply engrossed)			
Consistent image					
Trust			External stability (protects information)		Assurance (feel safe)
Response time	Download time		Internal stability (loading speed, reliability)	Access (responsive, loads quickly)	
Customer service*	Responsiveness (FAQs, feedback)		Order processing (prices good, delivered on time, exchanges); Comms Interface (user community is formed, fast accurate answers to questions)		Reliability (get what asked for); responsiveness (if problems); empathy
Not Included in WebQual				Hyperlinks	Asset specificity (prod avail from multiple sites)

Table A1. Constructs in Five Articles in June/September 2002 Special Issue of ISR

Notes: Agarwal and Venkatesh 2002 [3] not included here since only a single question on each construct or subconstruct. Did not do required tests of discriminant validity. * Customer Service not included in this version of WebQual.

Appendix 2. WebQual Items by Construct

Usefulness

Informational Fit-to-Task (INFO)

The information on the Web site is pretty much what I need to carry out my
    tasks.
The Web site adequately meets my information needs.
The information on the Web site is effective.

Tailored information (TAILOR)

The Web site allows me to interact with it to receive tailored information.
The Web site has interactive features, which help me accomplish my task.
I can interact with the Web site in order to get information tailored to my
    specific needs.

Trust (TRUST)

I feel safe in my transactions with the Web site.
I trust the Web site to keep my personal information safe.
I trust the Web site administrators will not misuse my personal information.

Response Time (RESP)

When I use the Web site there is very little waiting time between my actions
    and the Web site's response.
The Web site loads quickly.
The Web site takes long to load.

Ease of Use

Ease of Understanding (EUDSTD)

The display pages within the Web site are easy to read.
The text on the Web site is easy to read.
The Web site labels are easy to understand.

Intuitive Operations (INTUIT)

Learning to operate the Web site is easy for me.
It would be easy for me to become skillful at using the Web site.
I find the Web site easy to use.

Entertainment

Visual Appeal (VISUAL)

The Web site is visually pleasing.
The Web site displays visually pleasing design.
The Web site is visually appealing.

Innovativeness (INNOV)

The Web site is innovative.
The Web site design is innovative.
The Web site is creative.

Emotional Appeal (EMOTION)

I feel happy when I use the Web site.
I feel cheerful when I use the Web site.
I feel sociable when I use the Web site.

Complementary Relationship

Consistent Image (CONSIMG)

The Web site projects an image consistent with the company's image.
The Web site fits with my image of the company.
The Web site's image matches that of the company.

On-Line Completeness (OLCOMP)

The Web site allows transactions on-line
All my business with the company can be completed via the Web site.
Most all business processes can be completed via the Web site.

Relative Advantage (RELADV)

It is easier to use the Web site to complete my business with the company than it is to telephone, fax, or mail a representative.
The Web site is easier to use than calling an organizational representative agent on the phone.
The Web site is an alternative to calling customer service or sales.

Intent to Reuse the Web Site

For All Administrations Except 2002/2003:

How likely or unlikely would you be to make a purchase from this Web site?

How likely or unlikely would you be to revisit this Web site?

For 2002/2003 Administrations:

If I needed this product or service in the future, I would be likely to buy it from this Web site.

If I needed this product or service in the future, I would probably revisit this Web site.

If I needed this product or service in the future, I would probably try this Web site.

If I needed this product or service in the future, I would probably end up making a purchase from this Web site.

I would recommend this Web site to a friend interested in this product or service.

Appendix 3

Round 3 (2000 summer/fall)

	INFO	TAILOR	TRUST	RESP	EUDSTD	INTUIT	VISUAL	INNOV	EMOTION	CONSIMG	OLCOMP
INFO	0.75										
TAILOR	0.37	0.40									
TRUST	0.34	0.31	0.33								
RESP	0.56	0.56	0.29	0.33							
EUDSTD	0.61	0.62	0.24	0.28	0.67						
INTUIT	0.36	0.42	0.32	0.39	0.47	0.31					
VISUAL	0.50	0.55	0.38	0.35	0.46	0.36	0.70				
INNOV	0.46	0.47	0.35	0.36	0.40	0.38	0.52	0.60			
EMOTION	0.42	0.48	0.34	0.40	0.47	0.39	0.64	0.60	0.54		
CONSIMG	0.53	0.52	0.30	0.30	0.40	0.56	0.29	0.39	0.40	0.39	
OLCOMP											
RELADV	0.59	0.55	0.30	0.22	0.37	0.42	0.30	0.37	0.37	0.41	0.56

Round 4 (2002/2003)

	INFO	TAILOR	TRUST	RESP	EUDSTD	INTUIT	VISUAL	INNOV	EMOTION	CONSIMG	OLCOMP
INFO	0.76										
TAILOR	0.41	0.37									
TRUST	0.28	0.30	0.19								
RESP	0.60	0.59	0.32	0.45							
EUDSTD	0.66	0.61	0.35	0.41	0.74						
INTUIT	0.43	0.46	0.33	0.25	0.51	0.36					
VISUAL	0.50	0.52	0.37	0.27	0.48	0.38	0.70				
INNOV											
EMOTION	#	#	#	#	#	#	#	#	#		
CONSIMG	0.56	0.48	0.38	0.23	0.51	0.46	0.50	0.50	#	0.45	
OLCOMP	0.62	0.56	0.45	0.36	0.55	0.61	0.27	0.36	#	0.42	0.60
RELADV	0.71	0.55	0.30	0.35	0.55	0.71	0.23	0.32	#		

Table A3. Round 3 and 4 Correlations of Distinct Factors

Emotional appeal is not included for round 4 (2002/2003) because a different version of the emotion questions was used for that data collection.

ELEANOR T. LOIACONO (eloiacon@wpi.edu) is an assistant professor of MIS at Worcester Polytechnic Institute. She earned her Ph.D. in information systems at the University of Georgia and has published in *Communications of the ACM, IEEE Computer,* and *Journal of Information Technology and People.* Her research interests include IT accessibility, Web site quality, human-computer interaction, and affective issues in information systems.

RICHARD T. WATSON (rwatson@terry.uga.edu) is the J. Rex Fuqua Distinguished Chair for Internet Strategy and director of the Center for Information Systems Leadership, Terry College of Business, University of Georgia. He has published in leading journals in several fields and has authored books on data management and electronic commerce. His current research focuses primarily on electronic commerce and IS leadership. He has given invited seminars in more than 20 countries for companies and universities. He is a visiting professor at Agder University College, Norway, a consulting editor to John Wiley & Sons, and co-leader of the Global Text Project. He has been president of AIS, a co-chair of ICIS, and a senior editor for *MIS Quarterly.*

DALE L. GOODHUE (dgoodhue@terry.uga.edu) is a professor of MIS and the C. Herman and Mary Virginia Terry Chair of Business administration at the University of Georgia's Terry College of Business. He has published in *Management Science, MIS Quarterly, Decision Sciences, Sloan Management Review,* and other journals, and is an associate editor for *Management Science* and a former associate editor for *MIS Quarterly.* His research interests include measuring the impact of information systems, the impact of task-technology fit on individual performance, the management of data and other IS infrastructures/resources, and the impact of ERP systems on organizations.