

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/236346018>

WebQual™: A Web Site Quality Instrument

Conference Paper · January 2002

CITATIONS

98

READS

1,517

3 authors, including:



Eleanor T. Loiacono

Worcester Polytechnic Institute

66 PUBLICATIONS 902 CITATIONS

SEE PROFILE



Richard T. Watson

University of Georgia

232 PUBLICATIONS 9,225 CITATIONS

SEE PROFILE

WEBQUAL: A MEASURE OF WEBSITE QUALITY¹

Eleanor T. Loiacono, Worcester Polytechnic Institute, Worcester

Richard T. Watson, University of Georgia, Athens

Dale L. Goodhue, University of Georgia, Athens

ABSTRACT

The paper presents WebQual™, a Web site quality measure with 12 dimensions. Development was based on an extensive literature review, and interviews with Web designers and Web visitors. The instrument was refined using two successive samples, and the validity of the final instrument was tested with a third confirmatory sample.

INTRODUCTION

There does not exist a comprehensive instrument specifically designed to focus on the consumer's perception of Web site quality in the context of predicting the behavior of reuse of the site. This paper seeks to address that gap, utilizing as a general underlying model the Theory of Reasoned Action (TRA) (Ajzen et al. 1980; Fishbein et al. 1975), and particularly the TRA as applied to information technology utilization in the Technology Acceptance Model (TAM) (Davis et al., 1989).

The Theory of Reasoned Action (Ajzen et al. 1980; Fishbein et al. 1975) states that individuals' behavior (in this case revisiting or purchasing from a Web site) can be predicted from their intentions, which can be predicted from their attitudes about the behavior and subjective norms. Following the chain of prediction further back, attitudes can be predicted from an individual's beliefs about the consequences of the behavior. Subjective norms can be predicted by knowing whether significant other individuals think the behavior should or should not be done. TRA is a very general theory, and as such does not specify what specific beliefs would be pertinent in a particular situation.

Davis (1989) applied TRA to a class of behaviors that can be loosely defined as “using computer technologies,” and produced a Technology Acceptance Model (TAM). Davis argues that for the behavior of “using computer technologies,” two particular beliefs are predominant in predicting behavior: perceived ease of use and perceived usefulness. We also believe that there may be

¹ Loiacono, Eleanor T., Richard T. Watson, and Dale L. Goodhue. 2002. WebQual™: a measure of Web site quality. In *AMA Winter Conference*. Austin, TX.

multiple distinct dimensions of ease of use and of usefulness, as well as other categories of beliefs such as “entertainment” that together predict intentions to reuse a Web site. Determining the relevant specific dimensions of WebQual, and developing an effective instrument to measure those is the subject of the rest of the paper.

There are many frameworks for thinking about measurement validity. Bagozzi (1980) and Bagozzi and Phillips (1982) are used in this paper due to their comprehensive coverage of six key components of validity, which are explained along with the process used to develop WebQual.

INSTRUMENT DEVELOPMENT PROCESS

The goal was to develop a valid measure of Web site quality that would predict Web site reuse. The overall process included four stages, which together addressed each of Bagozzi’s validity concerns.

1. We move *beyond* and *within* the two constructs of ease of use and usefulness. That is, we examined whether there are other categories of beliefs that also need to be considered, and whether there are there distinct dimensions of “ease of use” and “usefulness” that should be considered separately.
2. We developed questions for each of the dimensions of WebQual identified in Stage 1. The initial result was an eighty-eight-item instrument that measured 13 distinct beliefs about a Web site.
3. The instrument was refined by administration to two different samples (N = 511 and N = 336). After each administration, the measurement validity of the constructs was analyzed and problem questions pruned, revised, or replaced and redundant dimensions collapsed. This resulted in an instrument with 36 questions measuring 12 dimensions of WebQual.
4. A confirmatory analysis of the overall measurement validity of the final instrument was conducted using a new sample of 307 subjects. The instrument demonstrated strong measurement validity for the four validity issues that can be empirically assessed.

STAGE 1: DEFINING THE DIMENSIONS OF WEBQUAL

In order to reveal the pertinent dimensions of Web site quality and establish content validity, a four-pronged effort was employed more or less simultaneously. First, a review of the MIS and marketing literature revealed existing constructs related to quality and customer satisfaction. In parallel with this effort, we conducted three exploratory research projects to ensure the comprehensiveness of the constructs. These included soliciting criteria from Web surfers, interviewing Web designers, and studying a large organization's standards for Web site design.

Initial dimensions of Web site quality

Five general categories of Web site quality arose from a literature review and exploratory research: ease of use, usefulness, entertainment, complementary relationship, and customer

service. These can be further broken down into 14 distinct dimensions of Web site quality as shown in Table 1.

----Table 1 about here----

STAGE 2: DEVELOPING THE ITEMS

Scale development can be either deductive or inductive (Hinkin 1998). We incorporated both approaches through an extensive literature review (inductive) and an exploratory research (deductive) phases.

An initial set of 142 candidate items was developed based on 13 constructs arising out of the literature review and our exploratory studies. This list of items was then refined based on an approach used by Davis' (1989) in his pretest of measures for the TAM. Mindful of the cognitive complexity of handling all 13 constructs (Miller 1956), we opted to reduce the difficulty of this initial screening. Twenty experienced Web users from a large southeastern University (5 graduate and 15 undergraduates students) rated the items on how well they corresponded to the four high-level categories of Web site quality (ease of use, usefulness, entertainment, and complementary relationships).

A non-statistical cluster analysis (similar to Davis, 1989) was performed by incorporating an item into one of the four high-level constructs if at least fifty percent (10 out of 20) of the subjects ranked the item as one of the top three for these particular constructs. This resulted in an initial WebQual instrument of 88 items covering a possible 13 constructs.

STAGE 3: REFINEMENT

In order to prevent item order bias, two random order versions of the initial instrument were created. To preventing inflating reliabilities from artificially high correlations where subject answered adjacent questions using anchoring and adjustment, items pertaining to a similar construct were separated from other items. Items were measured using a seven-point Likert scale. In addition, reverse scored items were included to ensure respondents were alert while completing the survey and to eliminate response bias (Hensley 1998; Spector 1992).

The instrument was refined by examining its reliability and discriminant validity after each of two distinct administrations. With 88 questions in the initial questionnaire and a rule of thumb for factor analysis of at least five times as many observations as there are variables to be analyzed (Hair et al. 1998), at least 440 subjects were required. Data were collected from 510 undergraduates in round 1.

Subjects were given a context (e.g., "Imagine it is your friend's birthday and you are searching for a good gift—a book.") and told to explore a designated Web site as if they were considering which book to buy for their friend, and then to complete the questionnaire. Three of each of four different types of Web sites were used (12 in all). The sites were chosen for their quality variability, based on rankings of specific Web sites generated by subjects in the exploratory research phase. In order to control for time of day bias, the time of day the sites were visited was varied.

Item assessment and purification: round 1

Data analysis and purification consisted of the following three steps. First, the Cronbach's alpha for each measure of the 13 target constructs was calculated. Items that were determined to decrease the reliability (alpha) of a construct's measure were deleted and the process continued until no item's removal increased a construct's overall alpha. The end result was the removal of eleven items.

As a second means to identify internal consistency problems, those items found to possess low correlations with similar traits (i.e., less than .40) were removed from the instrument. A total of 13 items were deleted during this phase, while one was simply modified in order to clarify its meaning.

The final step consisted of removing items that appeared to have discriminant validity problems. Items were removed if they correlated more highly with items measuring different constructs than they did with items in their intended construct (Campbell et al. 1959; Goodhue 1998). Under these criteria, eight items were deleted.

After the deletions, each construct was reviewed to ensure that at least five items per dimension remained. (This permitted us to drop up to two items for each dimension, if we discovered measurement validity problems, and still had at least three items per dimension.) For those dimensions that were underrepresented, additional items closely related to the remaining items were added. Twenty-seven items were added—resulting in an 83-item instrument.

STAGE 4. FINAL ITEM SELECTION AND ASSESSMENT OF MEASUREMENT

A second round of data collection allowed testing of the measurement validity of the second version of the instrument. Data were collected from 336 undergraduate students. A two-step process was employed to select the subset of the questions to be included in the final version of WebQual.

Discriminant validity: round 2

Discriminant validity for the second version of the questionnaire was first assessed using exploratory factor analysis (EFA). Five of the 13 constructs (information quality, fit-to-task, interactivity, innovativeness, and business processes) appeared to have some possible discriminant validity problems. All other constructs loaded on separate factors.

To explicitly test for discriminant validity of the five problematic constructs, we used confirmatory factor analysis (CFA) and chi-squared difference tests. The results revealed that the information quality and fit-to-task constructs should be combined. Thus, the outcome of discriminant validity analysis was to reduce the WebQual measure from 13 to 12 constructs.

Validity of final instrument

Once the 12 key concepts of WebQual were identified, the top three items loading on each factor were chosen for the final questionnaire. Since a construct should have at least three items

(Cronbach et al. 1955) and we sought a parsimonious instrument (lengthy questionnaires typically have a lower response rate) (Babbie 1998), three items were selected for each construct. This final version of 36 items was used to assess the empirically testable validity concerns (Bagozzi's last four concerns from Table 1).

Confirmatory factor analysis

A final confirmatory factor analysis was run using a sample of 307 undergraduate students. The results indicate strong support for the overall fit of the model. Both the RMSEA (0.060) and SRMR (0.053) are within conservative limits of acceptable error. Further the RNI (.92) and NNFI (.91) confirm good model fit.

Internal consistency (reliability)

The reliability of the final questionnaire (12 constructs with 3 questions each) was calculated using Cronbach's alpha (Cronbach 1951). The alphas of the twelve constructs ranged from .72 to .93, with 10 of the 12 constructs having an alpha greater than .80. Of the two remaining constructs, one had an alpha of .79, and one had an alpha of .72

Discriminant validity

As a final check of discriminant validity, we tested all possible pairs of the 12 remaining constructs to see if fit was improved when any pair was collapsed into a single construct. The two most highly correlated constructs are information fit to task and interactivity, correlated at .90. Even though these two are highly correlated, the discriminant analysis confirms that all 12 constructs are separate dimensions of a Web site's quality.

Convergent validity

Following the example of the development of SERVQUAL (Parasuraman et al. 1988), subjects were asked one additional question on overall Web site quality—*Overall, how would you rate the quality of this Web site?* (1 to 7 point scale anchored on “poor” and “excellent”). The total WebQual score (36 items) was computed and then compared to the overall quality of the website question. The two were correlated at .78 ($p < .001$) indicating convergent validity.

Nomological validity/predictive validity

Confidence in the measure increases, if it behaves as expected in relation to other acceptable constructs (Bagozzi 1979; Bagozzi 1980). In the case of WebQual, predictive validity is demonstrated by testing the ability of the instrument to accurately predict a Web visitor's intention to purchase from or revisit a Web site. The correlations between the composite WebQual measure and intention to purchase and intention to revisit were .56 and .53 respectively (both with $p < .0001$). This is good confirmation of nomological validity.

Adequacy of model fit

Four recommended fit indices (Vandenberg et al. 2000) indicate quite acceptable fit for the 12 construct, 36 question model (see Table 2). These indices provide consistent and reinforcing indications of the overall adequacy of WebQual.

----Table 2 about here----

DISCUSSION

This research makes two contributions. First, it provides practitioners and researchers with a validated reliable measure of Web site quality. The term “electronic commerce” may soon be redundant since nearly all commerce may be electronic (Porter, 2001). Even with slightly less optimistic growth of electronic commerce, firms will increasingly need a means of assessing the quality of a Web site. Web sites in many cases will fashion the customer’s view of the firm (Watson, 1998) and could have an important impact on performance. Second, this study adds to our understanding of TAM, a widely used MIS instrument, by revealing the components of ease of use and usefulness. Thus, it provides the basis for refining TAM to increase its diagnostic power.

LIMITATIONS

WebQual’s development was based on the responses of undergraduate business students to a selected group of Web sites. While these subjects are typical of a substantial body of Web users, they are not a representative sample of all users. Furthermore, many of the subjects were not on-going customers of the sites selected for assessment. 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. Further confirmatory research needs to be done with broad samples of on-going customers of a range of Web sites.

CONCLUSION

In the age of the Internet and electronic commerce, MIS and marketing need a means of assessing the effectiveness of a Web site. Our efforts have produced, we believe, a valid and reliable instrument for measuring Web site quality. WebQual should be able to support a range of important MIS and marketing studies as researchers attempt to understand what contributes to success in the electronic marketplace. The research presents the beginning of a cumulative research program by the authors, and we hope others, to understand the nature and characteristics of high quality Web sites.

REFERENCES

- Ajzen, I., and Fishbein, M. *Understanding attitudes and predicting social behavior* Prentice-Hall, Englewood Cliffs, NJ, 1980.
- Babbie, E. *The Practice of Social Research*, (8th ed.) Wadsworth, Belmont, CA, 1998.

- Bagozzi, R.P. "The role of measurement in theory construction and hypothesis testing: Toward a holistic model," in: *Conceptual and theoretical developments in marketing*, S.W.B. O.C. Ferrell, & C.W. Lamb, Jr. (ed.), American Marketing Association, Chicago, IL, 1979.
- Bagozzi, R.P. *Causal models in marketing* John Wiley, New York, 1980.
- Bagozzi, R.P., and Phillips, L. "Representing and Testing Organizational Theories: A Holistic Construal," *Administrative Science Quarterly* (27:3) 1982, pp 459-490.
- Campbell, D.T., and Fisk, D.W. "Convergent and discriminant validation by the multitrait-multimethod matrix," *Psychological Bulletin* (56) 1959, pp 81-105.
- Cronbach, L.J. "Coefficient alpha and the internal structure of tests," *Psychometrika* (16:3) 1951, pp 297-333.
- Cronbach, L.J., and Meehl, P.C. "Construct validity in psychological tests," *Psychological Bulletin* (52) 1955, pp 281-302.
- Davis, F.D. "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly* (13:3) 1989, pp 319-339.
- Fishbein, M., and Ajzen, I. *Beliefs, attitude, intention, and behavior: An introduction to theory and research* Addison-Wesley, Reading, MA, 1975.
- Goodhue, D. "Development and measurement validity of a task-technology fit instruments for user evaluations of information systems," *Decision Sciences* (29:1) 1998, pp 105-138.
- Hair, J.F., Jr., Anderson, R.E., Tatham, R.L., and Black, W.C. *Multivariate Data Analysis*, Upper Saddle River, NJ, 1998, p. 730.
- Hensley, R.L. "A review of operations management studies using scale development techniques," *Journal of Operations Management* (17) 1998, pp 343-356.
- Hinkin, T.R. "A brief tutorial on the development of measures for use in survey questionnaires," *Organizational Research Methods* (1:1) 1998, pp 104-121.
- Hoffman, D.L., and Novak, T.P. "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," *Journal of Marketing* (60:July) 1996, pp 50-68.
- Miller, G.A. "The magical number seven, plus or minus two: some limits on our capacity for processing information," *The Psychological Review* (63:2) 1956, pp 81-97.
- 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, pp 12-40.

Spector, P.E. *Summated rating scale construction: An introduction* Sage Publications Ltd., Newbury Park, 1992, p. 73.

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

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, pp 4-69.

Watson, R.T., Akselsen, S., and Pitt, L.F. "Attractors: building mountains in the flat landscape of the World Wide Web," *California Management Review* (40:2) 1998, pp 36-56.

Appendix: WebQual Items by Construct

USEFULNESS:
<i>Informational Fit-to-Task</i>
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.
<i>Interactivity</i>
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.
<i>Trust</i>
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.
<i>Response Time</i>
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:
<i>Ease of Understanding</i>
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.
<i>Intuitive Operations</i>
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:
<i>Visual Appeal</i>
The Web site is visually pleasing.
The Web site displays visually pleasing design.
The Web site is visually appealing.
<i>Innovativeness</i>
The Web site is innovative.
The Web site design is innovative.
The Web site is creative.
<i>Flow—Emotional Appeal</i>
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.
COMPLIMENTARY RELATIONSHIP:
<i>Consistent Image</i>
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.
<i>On-Line Completeness</i>
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.
<i>Better than Alternative Channels</i>
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 then calling an organizational representative agent on the phone.
The Web site is an alternative to calling customer service or sales.

Table 2: Overall Fit of the Full WebQual Model

	WebQual Round 2	WebQual Round 3	Recommended Cutoff
RMSEA	0.052	0.060	< 0.06 to 0.08
SRMR	0.047	0.053	< 0.06 to 0.08
RNI	.90	.92	> .90
NNFI	.94	.91	> .90

RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, RNI = Relative Noncentrality Index, NNFI = Non-normed Fit Index.

Table 1: Initial WebQual Dimensions

Higher Level Concept	Dimension	Description
Ease of Use	Ease of Understanding	Easy to read and understand.
	Intuitive Operation	Easy to operate and navigate.
Usefulness	Information Quality	The information provided is accurate, current, and relevant.
	Functional Fit-to-task	Meets task needs and improves performance.
	Interactivity	Tailored communication between consumers and the firm.
	Trust	Secure communication and observance of information privacy.
	Response Time	Time to get a response after a request or an interaction with a site.
Entertainment	Visual Appeal	The aesthetics of a Web site.
	Innovativeness	The creativity and uniqueness of site design.
	Flow	The emotional effect of using the Web site and intensity of involvement
Complementary Relationship	On-Line Completeness	Allowing all or most necessary transactions to be completed on-line (e.g., purchasing over the Web site)
	Better than Alternative Channels	Equivalent or better than other means of interacting with the company.
	Consistent Image	The Web site image is compatible with the image projected by the firm through other media
Customer Service	Customer Service	The response to customer inquiries, comments, and feedback.