Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2005 Proceedings

International Conference on Information Systems (ICIS)

December 2005

Content and Design Metrics for Web Sites: Combining Qualitative and Quantitative Results

Horst Treiblmaier
Vienna University of Economics and Business Administration

Andreas Pinterits
Vienna University of Economics and Business Administration

CONTENT AND DESIGN METRICS FOR WEB SITES: COMBINING QUALITATIVE AND QUANTITATIVE RESULTS

Horst Treiblmaier and Andreas Pinterits
Vienna University of Economics and Business Administration
Vienna, Austria
horst.treiblmaier@wu-wien.ac.at **andreas.pinterits@wu-wien.ac.at**

Abstract

Given the vital importance of the Internet as a communication and transaction channel, the evaluation of Web sites has become a central issue for both information systems researchers and practitioners. In contrast to traditional distribution and communication channels, which are usually adapted to different target groups, a company's Web site needs to appeal to a multitude of mostly anonymous visitors. Many different metrics have been developed to evaluate a Web site's quality as perceived by users. This paper explicitly focuses on users' perceptions of the content and design of Web sites and shows how a Delphi study as well as an exploratory factor analysis can be used to generate and cluster attributes of Web sites. Furthermore, different user groups are identified and clustered according to their preferences into the categories "Entertainment and Design," "Information," and "Individualization."

Keywords: Web site analysis, Web metrics, method triangulation

Introduction

The Internet has changed the way organizations interact with (prospective) customers. Especially in business-to-consumer relationships, a company's homepage has to appeal to a multitude of users. In addition to problems such as browser compatibility or download delay (Nah 2004), the question of adequate content and design emerges. Huizingh (2000, p. 123) refers to content as the information, features, or services that are offered in the Web site, while design is the way in which content is made available for Web visitors.

In the information systems literature, a variety of instruments have been developed in order to measure content-related and design-related aspects of Web sites, which are often considered to be direct antecedents of a Web site's success. Since in many cases no standardized definitions of the corresponding constructs exist, definitions provided by different authors may overlap to a certain extent (see Table 1).

Another approach uses certain features of Web sites, including corporate information, communication/customer support, presentation, and navigation, and classifies them into the categories of content and design (Robbins and Stylianou 2003). All of these research studies use pretested items (scales) or self-developed frameworks to assess the importance of the relevant constructs. This paper, meanwhile, seeks to measure consumers' perceptions of Web sites by identifying *attributes* well-suited for the subjective assessment of an entire Web site's content and design and aims to consolidate these attributes into factors. Unlike similar studies, we do not concentrate on parts of a site (e.g., the homepage) and do not assume a common understanding of the attributes (e.g., the site is interactive). Rather, we want to find out which attributes share a common understanding and may be used for general assessments of Web sites. By using and comparing qualitative and quantitative methods, we strive to assess whether both approaches are equally suited for building a measurement instrument.

Table 1. Constructs and Sources for Analyzing Web Sites	
Construct	**Supporting Literature**
Navigation/Organization	Palmer (2002)
(Perceived) Ease of Use	Abdinnour-Helm et al. (2005) based on Doll and Torkzadeh (1988); Agarwal and Venkatesh (2002); Palmer (2002); van Iwaarden et al. (2003); Zhang et al. (2000)
Usability	Agarwal and Venkatesh (2002); McKinney et al. (2002)
Navigation/Tangibles	van Iwaarden et al. (2003); Zhang et al. (2000)
Responsiveness/Specific Content	van Iwaarden et al. (2003); Palmer (2002); Aladwani et al. (2002)
Information/(Web) Content	Abdinnour-Helm et al. (2005) based on Doll and Torkzadeh (1988); Aladwani and Palvia (2002); González and Palacios (2004); Ho (1997); Palmer (2002); Yeung and Lu (2004)
Quality	McKnight et al. (2002); Zhang et al. (2000)
Promotion/Provision/Processing	Ho (1997)
Fun/Enjoyment/Entertainment/Delight	McKinney et al. (2002); Kim et al. (2002)
Layout/Presentation/Web Appearance/Convenience	Aladwani and Palvia (2002); Zhang et al. (2000); Kim et al. (2002)
Performance/Technical Adequacy/ Timeliness/Firmness	Abdinnour-Helm et al. (2005) based on Doll and Torkzadeh (1988); Aladwani and Palvia (2002); Kim et al. (2002)

This approach is similar to the methodology of Wang and Strong (1996), who created a framework of data quality by having data consumers assess the importance of a multitude of data quality attributes. Furthermore, this research methodology allows us to assess the psychological meaning (based on individual reactions) instead of the lexical meaning (based on conventions) that users associate with certain constructs (see Golden et al. 1989). Knowing these attributes helps both practitioners and researchers to understand how users actually perceive different constructs. In addition, the factors enable us to identify different user segments with divergent interests, expectations, and attitudes toward content and design.

Research Method

A total of three surveys were conducted to gather and merge attributes of Web sites. Due to the exploratory nature of this research, qualitative interviews seemed best suited for generating an extensive list of attributes (see phase 1 in Figure 1). In phase 2, the main focus was on generating a set of categories that best summarize the individual attributes. Therefore, a triangulation approach was chosen. A Delphi study (surveys a and b) and an online survey (survey c) were conducted. In the case of the Delphi study, the same set of experts who had originally generated the attributes placed them into different categories.

Afterward, in phase 3, the results of the previous two surveys b and c were compared and fitted together. This paper demonstrates how the results can be used to cluster user groups according to their individual attitudes.

The complementary combination of quantitative and qualitative research helps us to overcome ontological and epistemological problems and takes advantage of the benefits each of the two approaches provides. Although triangulation of methods has been strongly recommended for many years (e.g., Jick 1979; Webb et al. 1966), relatively few research projects have triangulated their methods. In this paper, we follow the approach suggested by Denzin (1978) and apply two different methods to one research problem to produce comparable data. This enables us to cross-validate our findings and improves the accuracy of the results of this study (Jick 1979).

Content validity is defined as the degree to which items in a measuring instrument represent the content universe to which the instrument will be generalized (Cronbach 1971). The experts were asked to generate as many items as possible, even if this led to redundancy, which was dealt with during later phases (see Torkzadeh and Dhillon 2002). Afterwards, in the second round, the experts had to group all items they found into categories. Only items considered representative by the majority of the experts were retained and thus constitute the result of the Delphi survey.

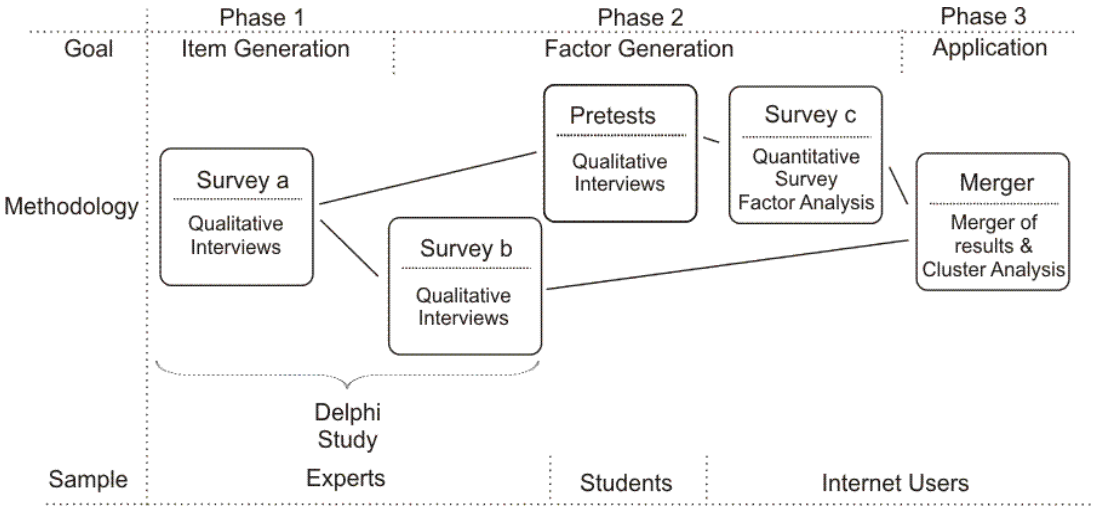

Figure 1. Research Design

External validity is determined by the scope of the quantitative survey. Straub et al. (2002) relate external validity to reasonable sample selection and size. We strived to gather a sample which at the best represents Austrian Internet users by using the customer base of two different companies. A total of 350 valid responses were collected in an online survey. We compared our sample to the universe of Austrian Internet users and found no bias toward various demographic characteristics (age, sex, education), which indicates a high level of external validity. In order to assess construct validity we used a procedure that resembles a multitrait–multimethod–matrix proposed by Campbell and Fiske (1959). By using two types of methods (qualitative interviews and a quantitative survey) and a variety of items grouped into different constructs, convergent and discriminant validity can be assessed by calculating the number of items assigned to the same category.

Since all surveys were conducted in the German language, the attributes were double-translated for the article (McGorry 2000). The items were first translated into English by the authors and translated back by an independent translator (see van der Heijden 2004). In a few cases, where no exact match between the original items and back-translation was found, another translator was consulted.

Delphi Study

A Delphi study can be used to obtain reliable consensus among experts by using a series of enquiries and controlled opinion feedback (Dalkey and Helmer 1963). We chose a design that fosters the generation and classification of ideas (Nambisan et al. 1999). As a first step, a panel of experts, capable of debating and judging the issues studied, had to be assembled. The knowledge and background required for this task was taken into account when selecting these experts (Malhotra et al. 1994). For the purpose of this survey, seven experts with varying backgrounds in information systems and information technology were chosen in order to incorporate as broad a range of ideas as possible into the study. In view of the exploratory nature of the survey, the first round was designed as a brain-storming phase. The experts were asked to come up with as many different attributes of Web sites as possible. For this purpose, we chose customer portal Web sites, which include informational content as well as a variety of entertainment and design features. This was intended to ensure that a broad range of attributes were generated and to avoid overlooking important aspects of users' perceptions of customer-oriented Web sites. When asking experts to name possible attributes for Web sites, care was taken to create a friendly, open atmosphere, so that the researchers had enough time to explain the specific goals of the survey to the experts and the latter had the possibility to clarify questions. All answers were pooled and formed the basis for our further research. As discussed above, the use of expert assessments not only provides a starting point for our further investigations but also strengthens content validity. This ensures that the operationalization of a construct adequately represents its domain of coverage (Straub 1989; Venkatraman and Grant 1986).

In the course of the first survey, the experts were asked to assess the classifications and to articulate several features of the services provided by the Web sites, which resulted in a total of 79 items. Following Moore and Benbasat (1991), the items were grouped into six categories ("Games and Fun," "Dynamics," "Information," "Individualization," "Emotion," and "Static Design Aspects").

Table 2. Results of the Delphi Study

Games, Fun and Dynamics	Information	Individualization	Emotion	Static Design Aspects
adventuresome	comprehensive	customized	dreamy	reputation enhancing
challenging	helpful	appropriate	emotional	artistic
quirky	informative	individualized	erotic	beautiful
entertaining	interesting	personalized	emotive	colorful
full of action	problem-orientated	personal	absorbing	creative
funny	relevant	target-group related	provocative	ornamented
humorous	supporting decisions	unerring	stimulating	elegant
interactive	up-to-date		touching	esthetic
dynamic				modern
performance measuring				multimedia-based
playful				plain
witty				prettifying
				professional
				self-explanatory
				stylish

The second survey was used to categorize the items collected. The experts were asked to sort the items into the category they considered most suitable (see Zhang von Dran 2000). The analysis of the results led to the consolidation of the two feature categories "Games and Fun" and "Dynamics" into a single category. Therefore, they were consolidated into one category named "Games, Fun, and Dynamics," which resulted in five remaining categories. Items that achieved at least five out of seven identical classifications were considered unequivocal and grouped into the respective categories.

The results of the Delphi study showed that the semantic meaning of several items was quite ambiguous. Therefore, 29 out of the 79 attributes with four or fewer matching classifications had to be eliminated (see Torkzadeh and Dhillon 2002). This resulted in a list of 50 features classified into five categories (see Table 2). For the following quantitative analysis, however, all original items were used in order to start both surveys with the same set of items to make for comparable results.

Online Survey

Based on the items generated from the qualitative interviews, an online survey was conducted. We used self-programmed sliders with a range from 1 to 100 to generate a magnitude scale (sometimes called a visual analogue scale, a graphical rating scale, or a continuous rating scale) instead of the commonly used Likert scales, thereby avoiding some weaknesses of the latter, for example, the loss of information due to the limited resolution of the categories (see Zeis et al. 2001) or the inadvertent influence of the investigator on the responses by constraining or expanding the response range available to the respondent (Treiblmaier et al. 2004). Previous research has shown that there are no overall differences between category scales and magnitude scales and that the latter can be considered a valid and reliable alternative, since both methods show considerable degrees of convergent and discriminant validity (Neibecker 1984). Although the loss of information from categorizing an unobserved, continuous variable into an ordered categorical scale can be reduced to a minimum when using at least five categories and multi-item scales (Srinivasan and Basu 1989), a multitude scale appeared to be the best research method available in view of the exploratory nature of our research design and the required eligibility of the data for the subsequent multivariate analyses.

Before starting the online survey, 10 pretests with students were conducted to eliminate items that were considered to be redundant or incomprehensible by the participants. The reduction in attributes was also necessary in order to eliminate unclear items that could lead to dropouts. The survey took place in June 2004 and was conducted in cooperation with Jowood, an Austrian producer of computer games, and AON Austria, the largest telecommunication portal in Austria with more than two million customers. An e-mail newsletter was sent out to customers of Jowood and an online banner was placed on AON's portal, both including a link to the survey. To offer incentives for survey participation, everyone who completed the survey was entered into a prize draw to win one of five Jowood computer games.

Users were asked to indicate how important they perceived each attribute to be for online customer portals. We did not give any example of certain sites in order to avoid bias, which may be caused by different levels of familiarity with a site. Furthermore, we have chosen consumer portals since they usually have to serve a multitude of heterogeneous user groups and therefore incorporate many of the attributes in which we were interested. After eliminating incomplete questionnaires and double counts, a total of 350 data sets remained. Since users were given the option to select "I don't know" if they found an attribute to be inappropriate, not all data sets could be used for further analyses. In general, most questionnaires were filled out completely. Only a single attribute (performance measuring) exhibited more than 5 percent of missing values.

In order to account for bias, we compared several demographic attributes of our sample (age, education, sex) with those of general Austrian Internet users (ORF 2005). While the age structure was quite comparable, we found a slight above-average percentage of users having graduated from a technical school or having completed an apprenticeship and furthermore a slight above-average percentage of males in our sample. A number of χ^2 tests revealed that none of these differences are statistically significant.

We used a principal axis factoring with promax as the method of rotation and listwise deletion. In contrast to the most commonly used principal component analysis, the results from principal axis factoring are more accurate in representing the population loadings (Widaman 1993). Furthermore, this research attempts to understand the latent structure of a set of variables instead of simply reducing them without interpreting the resulting variables in terms of constructs (see Conway and Huffcutt 2003. An oblique rotation is chosen instead of an orthogonal rotation since a correlation between the constructs is expected. Fabrigar et al. (1999, p. 287) state that besides getting "cleaner" solutions by using oblique rotation, simply "relying on an orthogonal rotation would also forfeit any knowledge of the existing correlations among factors." In all cases the number of factors was determined by using the scree test (see Velicer and Jackson 1990).

A KMO value of .883 indicates that the data are well-suited for factor analysis. To measure the reliability of the constructs, Cronbach's alpha was computed. The alpha values of "Entertainment and Design" (.92) and "Information" (.92) indicate that the set of items representing these constructs is sufficiently reliable, as opposed to the alpha value of "Individualization" (.54). In order to test the factor stability, we reran the analysis using different numbers of factors. The three-factor solution shown in Table 3 turned out to be the best in terms of interpretability. Seven items were removed since they failed to clearly load on a single factor. A number of items show factor loadings smaller than .5, which is a commonly accepted threshold for practical significance (Hair et al. 1998). In total, the three factors account for 39 percent of the variance in the variables. The results from the qualitative and the quantitative studies were compared in a next step.

Comparison of the Delphi Study and the Online Survey

In both surveys, referenced as survey b and survey c in Figure 1, we used the 79 attributes from the qualitative study as a starting point. The sorting process performed by the group of experts produced five factors, with a total of 50 items. The remaining 29 items were removed, since the experts considered them to be ambiguous or redundant. The items in our online survey design were reduced during the course of pretests which led to the elimination of 38 items. A factor analysis based on the data of the online survey helped to detect the underlying structure and was then used to extract three factors. Seven items were removed due to ambiguous loadings, which left us with a total of 34 items from the factor analysis. In a final step, the results from both research designs were combined. To make the results comparable, we had to reduce both item-sets to a joint basis, which led to the elimination of 22 items of the qualitative and 6 items of the quantitative set (see Figure 2).

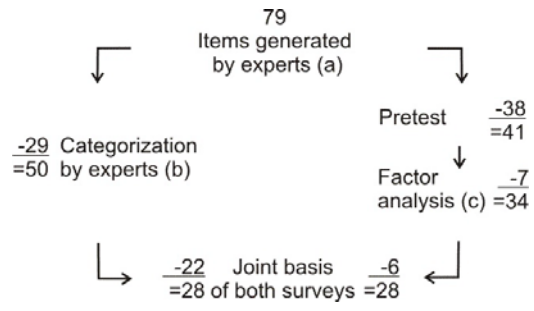

Figure 2. Elimination of Items

Table 3. Results of the Factor Analysis (n = 284)					
		F1	***F2***	***F3***	***Alpha***
Entertainment and Design	entertaining	.774			.92
	playful	.725			
	humorous	.724			
	funny	.706			
	colorful	.692			
	tensing	.670			
	quirky	.658			
	creative	.642			
	surprising	.623			
	full of action	.619			
	animated	.613			
	absorbing	.591			
	emotional	.572			
	challenging	.542			
	multimedia-based	.541			
	motivating	.525			
	inspiring	.507			
	beautiful	.504			
	modern	.465			
	provocative	.406			
Information	informative		.725		.84
	interesting		.695		
	professional		.668		
	comprehensive		.635		
	helpful		.625		
	self-explanatory		.616		
	sophisticated		.581		
	relevant		.578		
	problem-oriented		.488		
	up-to-date		.475		
	supporting decisions		.419		
Individualization	performance measuring			.536	.54
	individualized			.484	
	target-group-related			.451	
Variance Explained		.22	.12	.03	

Table 4. Comparison of the Qualitative and the Quantitative Survey					
		Entertainment and Design	**Information**	**Individualization**	**Total**
Delphi Study	Games Fun Dynamics	7		1	8
	Emotion	3	2	2	3
	Static Design Aspects	5	8	2	7
	Information	15	8	1	8
	Individualization	0	2		2
	Fitting				25
	Non-Fitting				3
	Total	15	10	3	28

Table 5. Attributes for Analyzing Web Sites		
Entertainment and Design	**Information**	**Individualization**
absorbing	comprehensive	individualized
beautiful	helpful	target-group related
challenging	informative	
colorful	interesting	
creative	problem-oriented	
emotional	relevant	
entertaining	supporting decisions	
full of action	up-to-date	
funny		
humorous		
modern		
multimedia-based		
playful		
provocative		
quirky		

Since the result of the Delphi study showed five categories while the factor analysis resulted in three, we mapped "Games, Fun and Dynamics," "Emotion, and "Static Design Aspects" to ""Entertainment and Design." A comparison of the different research designs can be seen in Table 4.

Of all 79 attributes being generated during the brain-storming phase by the experts, a total of 28 items emerged from both surveys, with 25 (89 percent) of these were assigned to the same categories, which indicates a high level of content validity. The result is an empirically derived list of attributes categorized into three constructs, "Entertainment and Design" (15), "Information" (8), and "Individualization" (2) (see Table 5). The construct named "Design and Entertainment" can be seen as representing the visual and hedonic aspect of a Web site, while "Information" represents the utilitarian point of view (Huang 2003).

Those attributes that were included in both research designs suggest a high level of congruency. Although information systems experts and students (who selected items in the pretests) eliminated items at different stages of the survey, the remaining attributes were assigned the same semantic meaning, either consciously (experts) or by analyzing the underlying data structures with the help of factor analysis (online survey).

To demonstrate the practical relevance of the findings, the next section shows how different user groups can be identified based on how important they perceive various content-related and design-related metrics.

Application

Based on the results from the Delphi study and the factor analysis, we used the data from the online survey to conduct a second factor analysis with the 25 attributes that were assigned to the same factors in both the qualitative and the quantitative study. Following the recommendations from Bensaou and Venkatraman (1995) and Punj and Stewart (1983), we conducted a cluster analysis: (1) standardized values were used for each variable, (2) the Euclidean distance was chosen as the similarity measure, and (3) Ward's minimum variance method was chosen to form clusters.

Since the clustering was based on standardized factors, no *absolute* comparison of the three influencing factors can be made. Generally speaking, "Information" was perceived as the most important variable with item medians ranging from 68 to 100. The attributes subsumed under "Individualization" had medians varying from 50 to 73, while "Entertainment and Design" attributes had the broadest range, from 20 to 73. Table 6 shows variable means and standard deviations for each of the three cluster groups. In addition, it gives the F-values and the significance levels associated with the test of equality of variable means across the three cluster groups.

Table 6. Cluster Analysis (n = 297)				
	Mean (S.D.) of Cluster Groups			
Variables	**Group 1 (n = 111)**	**Group 2 (n = 106)**	**Group 3 (n = 80)**	**F**
Entertainment and Design	-.040 (.678)	.791 (.735)	-.993 (.500)	168.2***
Information	-.772 (.920)	.538 (.484)	.359 (.483)	115.7***
Individualization	-.497 (.715)	.278 (.915)	.322 (.522)	39.0***

***$p < 0.001$

Group 1 attaches the least importance on "Information" and "Individualization" while sticking to the middle regarding "Entertainment and Design." Group 2 scores highest on "Entertainment and Design" and "Information" and also values "Individualization," while group 3 seemingly places more value on "Information" and "Individualization" than on "Entertainment and Design."

Furthermore, we tested for differences between the groups according to age, sex, education, occupation, and Internet usage. A one-way analysis of variance (ANOVA) revealed significant differences between the groups according to their age ($df = 2$, $F = 3.70$, $p < .05$). A Tukey HSD test revealed that the difference was caused by group 1 (average age: 33.72) and group 2 (average age: 39.97). No significant differences were found between the groups according to education and occupation, as was the case with Internet usage. However, significant differences according to sex ($\chi^2 = 5.99$, $df = 2$, $p = .05$) were detected, with, on the average, more women being in group 2 than in either of the other groups.

Conclusions and Further Research

To improve the measurement of Web metrics, we need to understand how people perceive the meaning of the respective constructs. This paper shows how different constructs can be built by using isolated attributes and by asking users to define how important they consider them. Merging qualitative and quantitative approaches to triangulate methods has helped to ensure the validity of the results. Interestingly, while the process of *selecting* attributes has led to different outcomes, the results of the *grouping* mechanisms turned out to be quite similar.

The resultant attributes may be used to assess different dimensions of Web sites and may also serve as a basis for further research. This can be done by using them as semantic differential measures, where users can rate Web sites on bipolar scales with contrasting adjectives on each end. Further research is needed to assess how the respective items are actually perceived by the users. The comparatively low levels of validity and reliability, which the factor "Individualization" exposes, might be seen as an indicator for an ambiguous comprehension of the respective items.

While this paper only sought to identify attributes that may be used to describe different dimensions of Web sites, the absolute importance of these features is not explicitly assessed. Previous research has demonstrated different ways to weight criteria, for example, based on the characteristics and the needs of the target markets (Evans and King 1999), by having evaluators determine the relative importance of a category (Agarwal and Venkatesh 2002), or by using methods such as analytic hierarchy process (AHP) (Moustakis et al. 2004). By bringing together these fields of research, new metrics can be developed.

We used a cluster analysis, based on the data from our online survey to show how different user groups can be identified on the basis of their relative preferences. More research is needed to learn how users evaluate individual Web sites. In addition, comparisons with existing Web metrics might be useful to determine how the general attributes identified in this study correlate with the items used in numerous other Web surveys.

Acknowledgments

The authors thank Sonja Brachowicz and Michael Swoboda from AON Austria and Michael Paeck from Jowood for their support in conducting the quantitative survey. We also thank Kerstin Kaim and Irene Pollach for their valuable aid at translation and retranslation of the items.

References

Abdinnour-Helm, S. F., Chaparro, B. S., and Farmer, S. M. "Using the End-User Computing Satisfaction (EUCS) Instrument to Measure Satisfaction with a Web Site," *Decision Sciences* (36:2), 2005, pp. 341-364.

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

Aladwani, A. M., and Palvia, P. C. "Developing and Validating an Instrument for Measuring User-Perceived Web Quality," *Information & Management* (39:6), 2002, pp. 467-476.

Bensaou, M., and Venkatraman, N. "Configurations of Interorganizational Relationships: A Comparison between U.S. and Japanese Automakers," *Management Science* (41:9), September 1995, pp. 1471-1492.

Campbell, D. T., and Fiske, D. W. "Convergent and Discriminant Validation by the Multitrait-Multimethod-Matrix," *Psychological Bulletin* (56), 1959, pp. 81-105.

Conway, J. M., and Huffcutt, A. I. "A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research," *Organizational Research Methods* (6:2), 2003, pp. 147-168.

Cronbach, L. J. "Test Validation," in *Educational Measurement*, R. L. Thorndike (Ed.), American Council on Education, Washington, DC, 1971, pp. 443-507.

Dalkey, N., and Helmer, O. "An Experimental Application of the Delphi Method to the Use of Experts," *Management Science* (9:3), 1963, pp. 458-467.

Denzin, N. K. *The Research Act*, McGraw-Hill, New York, 1978.

Doll, W. J., and Torkzadeh, G. "The Measurement of End-User Computing Satisfaction," *MIS Quarterly* (12:2), 1988, pp. 258-274.

Evans, J. R., and King, V. E. "Business-to-Business Marketing and the World Wide Web: Planning, Managing and Assessing Web Sites," *Industrial Marketing Management* (28:4), 1999, pp. 343-358.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J. "Evaluating the Use of Exploratory Factor Analysis in Psychological Research," *Psychological Methods* (4:3), 1999, pp. 272-299.

Golden, L. L., Alpert, M. I., and Betak, J. F. "Psychological Meaning: Empirical Directions for Identification and Strategy Development," *Psychology & Marketing* (6:1), Spring 1989, pp. 33-50.

González, M. F. J., and Palacios, B. T. M. "Quantitative Evaluation of Commercial Web Sites: An Empirical Study of Spanish Firms," *International Journal of Information Management* (24:4), 2004, pp. 313-328.

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

Ho, J. "Evaluating the World Wide Web: A Global Study of Commercial Sites," *Journal of Computer-Mediated Communication* (3:1), 1997.

Huang, M. -H. "Designing Website Attributes to Induce Experiential Encounters," *Computers in Human Behavior* (19:4), 2003, pp. 425-442.

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

Jick, T. D. "Mixing Qualitative and Quantitative Methods: Triangulation in Action," *Administrative Science Quarterly* (24:4), 1979, pp. 602-611.

Kim, J., Lee, J., Han, K., and Lee, M. "Businesses as Buildings: Metrics for the Architectural Quality of Internet Businesses," *Information Systems Research* (13:3), 2002, pp. 239-254.

Malhotra, M. K., Steele, D. C., and Grover, V. "Important Strategic and Tactical Manufacturing Issues in the 1990s," *Decision Sciences* (25:2), March/April 1994, pp. 189-214.

McGorry, S. Y., "Measurement in a Cross-Cultural Environment: Survey Translation Issues," *Qualitative Market Research* (3:2), 2000, pp. 74-81.

McKinney, V., Yoon, K., and Zahedi, F. "The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach," *Information Systems Research* (13:3), September 2002, pp. 296-315.

McKnight, H. D., Choudhury, V., and Kacmar, C. "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology," *Information Systems Research* (13:3), September 2002, pp. 334-359.

Moore, G. C., and Benbasat, I. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research* (2:3), 1991, pp. 192-222.

Moustakis, V. S., Litos, C., Dalivigas, A., and Tsironis, L. "Website Quality Assessment Criteria," in *Proceedings of the International Conference on Information Quality*, S. Chengular-Smith, J. Long, L. Raschid, and C. Seko (Eds.), Cambridge, MA, November 5-7, 2004, pp. 59-73.

Nah, F. F. -H. "A Study on Tolerable Waiting Time: How Long are Web Users Willing to Wait?," *Behaviour & Information Technology* (23:3), 2004, pp. 153-163.

Nambisan, S., Agarwal, R., and Tanniru, M. "Organizational Mechanisms for Enhancing User Innovation in Information Technology," *MIS Quarterly* (23:3), September 1999, pp. 365-395.

Neibecker, B. "The Validity of Computer-Controlled Magnitude Scaling to Measure Emotional Impact of Stimuli," *Journal of Marketing Research* (21:3), 1984, pp. 325-332.

ORF. *Austrian Internet Monitor* [survey online]. ORF Medienforschung, Vienna, Austria, 2005 (available online at **http://mediaresearch.orf.at/index2.htm?internet/internet_aim.htm**; accessed September 1, 2005).

Palmer, J. W. "Web Site Usability, Design, and Performance Metrics," *Information Systems Research* (13:2), 2002, pp. 151-167.

Punj, G., and Stewart, D. W. "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research* (20:2), May 1983, pp. 134-148.

Robbins, S. S., and Stylianou, A. C. "Global Corporate Web Sites: An Empirical Investigation of Content and Design," *Information & Management* (40:3), 2003, pp. 205-212.

Srinivasan, V., and Basu, A. K. "The Metric Quality of Ordered Categorical Data," *Marketing Science* (8:3), 1989, pp. 205-230.

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

Straub, D. W., Hoffman, D. L., Weber, B. W., and Steinfeld, C. "Toward New Metrics for Net-Enhanced Organizations," *Information Systems Research* (13:3), 2002, pp. 227-238.

Torkzadeh, G., and Dhillon, G. "Measuring Factors that Influence the Success of Internet Commerce," *Information Systems Research* (13:2), June 2002, pp. 187-204.

Treiblmaier, H., Pinterits, A., and Floh, A. "Antecedents of the Adoption of E-Payment Services in the Public Sector," in *Proceedings of the 25th International Conference on Information Systems*, R. Agarwal, L. Kirsch, and J. I. DeGross (Eds), Washington, DC, 2004, pp. 65-76.

van der Heijden, H. "User Acceptance of Hedonic Information Systems," *MIS Quarterly* (28:4), 2004, pp. 695-704.

van Iwaarden, J., van der Wiele, T., Ball, L., and Millen, R. "Applying SERVQUAL to Web Sites: an Exploratory Study," *The International Journal of Quality & Reliability Management* (20:8/9), 2003, pp. 919-935.

Velicer, W. F., and Jackson, D. N. "Component Analysis versus Common Factor Analysis: Some Issues in Selecting an Appropriate Procedure," *Multivariate Behavioral Research* (25:1), 1990, pp. 1-28.

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

Wang, R. Y., and Strong, D. M. "Beyond Accuracy: What Data Quality means to Data Consumers," *Journal of Management Information Systems* (12:4), 1996, pp. 5-33.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. *Unobtrusive Measures: Non-Reactive Research in the Social Sciences*, Rand McNally, Chicago, 1966.

Widaman, K. F. "Common Factor Analysis Versus Principal Component Analysis: Differential Bias in Representing Model Parameters?," *Multivariate Behavioral Research* (28:3), 1993, pp. 263-311.

Yeung, L. W., and Lu, M. -T. "Gaining Competitive Advantages through a Functionality Grid for Website Evaluation," *Journal of Computer Information Systems* (44:4), 2004, pp. 67-77.

Zeis, C., Regassa, H., Shah, A., and Ahmadian, A. "Goodness-of-Fit Tests for Rating Scale Data: Applying the Minimum Chi-Square Method," *Journal of Economic and Social Measurement* (27:1/2), 2001, pp. 25-39.

Zhang, P., and von Dran, G. M. "Satisfiers and Dissatisfiers: A Two-Factor Model for Website Design and Evaluation," *Journal of the American Society for Information Science* (51:14), 2000, pp. 1253-1268.

Zhang, X., Keeling, K. B., and Pavur, R. J. "Information Quality of Commercial Web Site Home Pages: An Exploratory Analysis," in *Proceedings of the 21st International Conference on Information Systems*, W. J. Orlikowski, S. Ang P. Weill H. C. Krcmar, and J. I. DeGross (Eds), Brisbane, Australia, 2000, pp. 164-175.