

WEBSITE QUALITY ASSESSMENT CRITERIA

(Research paper: IQ Concepts, Tools, Metrics, Measures and Methodologies)

Vassilis S. Moustakis^{1,2}, Charalambos Litos¹, Andreas Dalivigas¹, and Loukas Tsironis¹

¹ Department of Production and Management Engineering, Technical University of Crete, Chania 73100, Greece

² Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), Science and Technology Park of Crete, Heraklion 71110, Greece

moustaki@logistics.tuc.gr

Abstract: The article presents a hierarchical framework, which supports website quality assessment. The framework is composed of a hierarchical structure of criteria and sub-criteria and makes use of the Analytic Hierarchy Process to enhance criterion and sub-criterion weight value assessment. To validate the framework we conducted an experiment, which involved the assessment of the websites of the three cellular phone service providers in Greece by 122 users. Results confirmed framework validity and statistical factor analysis supported reduction of the original model to a website quality scoring framework, which involves nine composite criteria.

Key words: Website quality assessment, Analytic Hierarchy Process.

INTRODUCTION

Websites are part of our everyday life and are used to exchange and to convey information between user communities. Conveyed information comes in different types, languages and forms and incorporates text, images, sound, and video intended to inform, persuade, sell, present a viewpoint or even change an attitude or belief. Despite website proliferation, assessment of site quality remains a challenging area of research [15]. Quality relates to customer satisfaction and also with the level of accomplishment of user expectation when interfacing a website [11, 28]. International Standards Organization (ISO) standards 14598-3 [16] and 9126 [17] incorporate models, which focus on general software external characteristics that must be accomplished when the software is in use. The two standards capture external characteristics, but fail to account the internal characteristics that arise during the creation process [21].

As employed in this article, quality captures perceptual aspects, likely to be involved in human – website interaction. These aspects focus on the affective and cognitive royalty of a site, are qualitative and are subjectively assessed by the user community. Such focus on quality maps to customer satisfaction assessment [7] and contributes to the emergence of website quality as an aggregate composite that brings together formal metrics and perceptual user traits.

Scientific literature identifies several aspects, or criteria, of quality, which are often aggregated to [1, 2, 5, 14, 20, 23, 24, 28, 31, 32]:

1. Content
2. Navigation
3. Design and structure
4. Appearance and multimedia
5. Uniqueness

We used the above criteria as a starting point. These criteria were further decomposed to sub-criteria to formulate a model able to capture user's perception about quality in website use. Quantitative assessment about criterion and sub-criterion weight values was motivated via the use of the Analytic Hierarchy Process, or AHP for short [25]. Then we applied the criterion / sub-criterion hierarchy and assessed website quality of the three cellular phone providers in Greece, namely: Cosmote [6], Telestet [27] and Vodafone [30]. Assessment of website quality was achieved via the judgment elicitation by 122 site users. Users were first asked to assess criterion and sub-criterion weight values and then to use these values to assess quality preference about each website. During assessment identity of companies was hidden and companies were referred as A, B and C – tagging was random with respect to the three companies. Orientation was not towards the assessment of quality of the specific company websites; the selected websites were used to enable and facilitate quality criteria definition and validation. User judgments, both on criteria and the selected websites formed the database upon which AHP and statistical and factor analysis were applied to study criterion significance and synergy between criteria. However, research orientation was towards criteria definition, criterion weight value assessment and ultimately quality model validation.

Unique aspects of research reported herein include the homogeneity of sites and of the users who participated in site assessment. Both aspects of homogeneity mark a major difference from other surveys and aim to contribute to the specification of site quality characteristics or site design characteristics, which are based on actual website use; Ivory et al [18] report that such studies are missing from the literature.

In the sections that follow, we overview the method and criteria used, present research methodology, briefly overview AHP, and present and discuss findings from AHP application, statistical and factor analysis. We conclude the article by discussing areas for further research in the future.

BACKGROUND

To enhance quality assessment we generated a criterion / sub-criterion three echelon hierarchy, which is presented in Figure 1. The top level identifies the scope of the research, e.g., website quality. The second level captures the five main criteria each of which is further decomposed to a set of sub-criteria that range from three to seven. Criterion and sub-criterion definition was based on literature review and was oriented towards the formulation of a rich set of criteria and sub-criteria, which would enable reduction and aggregation, based on formal data and rotated factor analysis. The bottom level includes the three companies whose websites are compared using the criteria structure. Companies are compared with respect to each sub-criterion and results are synthesized leading to the computation of a total score for each company.

Criteria and sub-criteria are not independent. Altogether they form an initial space, which is oriented towards the evolution of a quality assessment tool. The reader should, however, notice that the characteristics that contribute to website quality are interdependent and thus the criterion and sub-criterion definitions carried through the hierarchy of Figure 1 are also interdependent. We have selected to work with a rather large set of criteria to allow users to find what they would like to see in assessment.

We used statistical factor analysis to clear up the hierarchy and to suggest a handful of criteria. Criterion and sub-criterion definitions may linguistically differ from other studies. Nevertheless, the field lacks a standardized lexicon. Criterion and sub-criterion definitions and hierarchical structure focus more on website use semantics rather than on page level metrics. Ivory et al [18] focus on website page level metrics to distinguish between “good” and “not so good” web pages using the webby 2000 award results on sites selected from six topical categories suggested by [28]. The criteria space used in this article partially expands the list of criteria used by the International Academy of Digital Arts and Sciences in webby awards [28]. The Academy bases its webby awards on six criteria, namely: (1) content, (2) structure and navigation, (3) visual design, (4) functionality, (5) interactivity, and (6) overall impression. Our study focuses on, and expands the first four Academy criteria. It handles overall impression about quality, based on the analytic rating judgments on the selected focus criteria. In the subsections that follow we discuss each main criterion and corresponding sub-criteria.

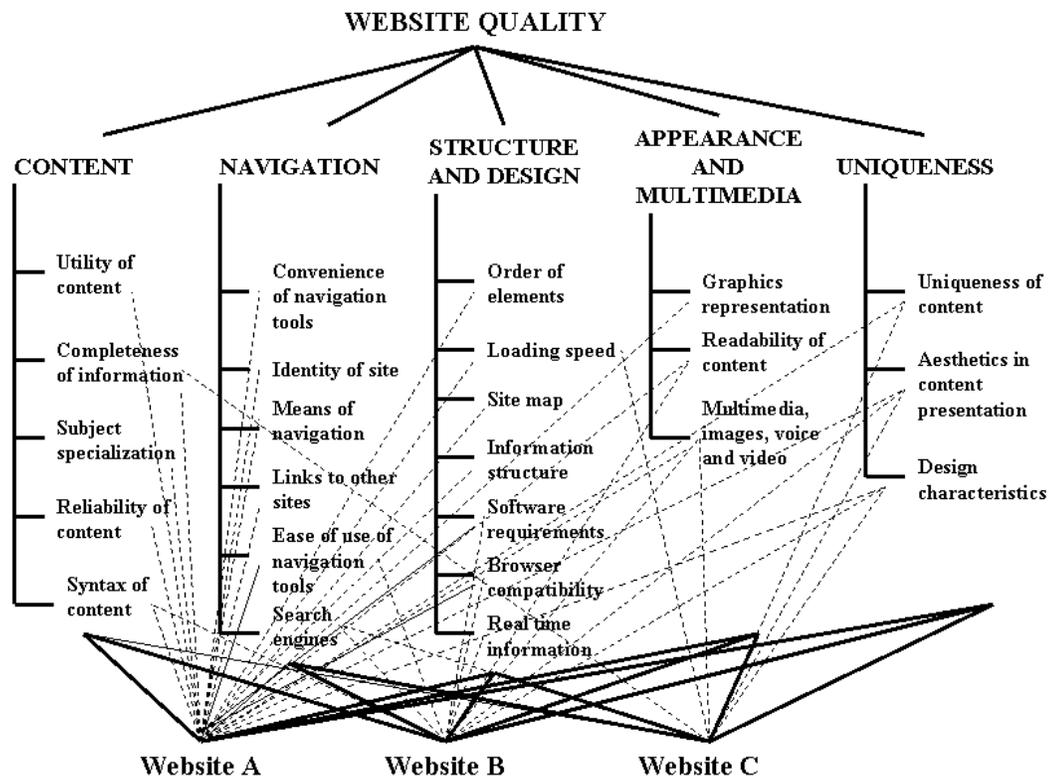


Figure 1: Criterion and sub-criterion assessment space. Top level includes the main objective: website quality. The second level captures the five main criteria and the third level captures the sub-criteria. Companies are listed at the bottom level of the hierarchy. Preference about the website of each company is calculated via pair-wise assessment with respect to each sub-criterion (this is denoted via the dotted lines). The solid lines that link the company website with the criteria indicate that all companies are assessed with respect to all sub-criteria and to exemplify the process we have linked using dotted lines website A with all sub-criteria and websites B and C with few sub-criteria.

Content

Is information conveyed by the site reliable and error-free? Content reflects quality, completeness, degree of specialization or generalization and reliability of information included in the website. Content also relates to the responsiveness of a website to satisfy a user inquiry and to the trustfulness about the information, which is included in the site [3, 12]. The specific sub-criteria that capture trust are utility of content and reliability of content (Table 1). Diligence, comprehension, completeness and language of information provided to the user make up the dimensions of content [31]. Finally, reliability implies that content should be modified, corrected and improved continuously to reflect upon environmental changes [16, 17]. Dimensions, or sub-criteria, of content are summarized in Table 1.

Sub-criterion	Definition
<i>Utility of content</i>	Captures the degree to which website incorporates essential, useful, trustful and up to date information: "... all pages should state the date on which the page was last updated..." [13].
<i>Completeness of information</i>	Captures website's explanatory profile with respect to the information contained within the site: "...information should be presented in a directly usable format that does not require decoding, interpretation, or calculation..." [13]
<i>Subject specialization</i>	Captures the degree to which website offers specific information to those who need it. Many researchers argue that websites should: (a) contain material tuned to satisfy specific needs versus including material that is targeted to general audiences, (b) organize information hierarchically, with more general information appearing before more specific detail; and (c) allow the user to delve as deeply as needed, but to stop whenever sufficient information has been received [23].
<i>Reliability of content</i>	Captures user's perception with respect to correctness and trustworthiness of information conveyed by the site.
<i>Syntax of content</i>	Captures means of content presentation including text, image, voice or graphic data.

Table 1: Sub-criterion dimensions of **content**. Dimensions form the sub-criteria identified in Figure 1.

Navigation

Where am I? How can I get to other places? Where can I go? Or, are directions for using the site provided? [19]. Navigation reflects the support provided to the user when moving in and around the site. Elements of navigation include: easiness of moving around, easiness in understanding site structure, and availability and validity of links. For instance, increasing the number of site links does not necessarily contribute to adding value to the site. Furthermore, links should be periodically reviewed and "linking mania" be avoided [4]. Dimensions, or sub-criteria, of navigation are summarized in Table 2.

Structure and Design

Does website elements sequence presentation reflect the importance, priority and frequency with which information should be accessed and used [13]? Are information categories clear to the users that are intended for [31]? Does site use require a browser that is too advanced? Structure and Design incorporates aspects that affect order of presentation, speed and browser. For example, existence of a site map enhances website value because it supports differentiation between information categories included in the site; in addition, if colors are used each category should be assigned to a unique variant [31]. Furthermore, "lighter" sites may prove easier to use since they tend to be more compatible with alternative software environment. Other aspects of structure and design include, but not limited to, background (use of complex background schemes tends to degrade site usability) and togetherness of information:

functionally related data should be displayed together, on one page [13]. Dimensions, or sub-criteria, of structure and design are summarized in Table 3.

Sub-criterion	Definition
<i>Convenience of navigation tools</i>	Captures easiness in surfing around the site. For instance, labels should be placed in proximity to their related data fields, or, users should always be given the chance to return to “home page” [31].
<i>Identity of site</i>	Reflects uniqueness of the site and the characteristics that make the site unique in a world full of sites.
<i>Means of navigation</i>	Reflects the availability of tools that support navigation in and around the site, such as labels, buttons, etc.
<i>Links to other sites</i>	Captures the informative perception of connections to other sites or data repositories that the website gives to the visitor.
<i>Ease of use of navigation tools</i>	Captures broader “easiness” issues related with navigation.
<i>Search engines</i>	Captures both availability and readiness of search engines embodied in the site.

Table 2: Sub-criterion dimensions of **navigation**. Dimensions form the sub-criteria identified in Figure 1.

Sub-criterion	Definition
<i>Order of elements</i>	Reflects information presentation consistency.
<i>Loading speed</i>	Reflects website’s loading speed. Loading speed may vary according to software platform and network speed. However, sites that rely heavily on image, voice, or video data tend to be slower compared with sites, which are simpler. Specifically as far as it concerns mobile www site developers, “Nokia research centre found that quick download time is more important than visual look when it comes to subjective satisfaction of the users” [29].
<i>Site map</i>	Reflects quality (or even availability) of site map. The site map is considered as part of structure and design because its existence and effectiveness relates to the website layout and organization and is independent of content.
<i>Information structure</i>	Reflects order and togetherness of information included in the website.
<i>Software requirements</i>	Reflects necessity of highly specialized software to access and to use the site – this is a negative criterion (more specialized software is expected to degrade user perception about a site).
<i>Browser compatibility</i>	Reflects to the ability to access and to use the site using a variety of different browsers. Websites should be designed for browsers at least one version lower than the most current version [31]
<i>Real time information</i>	Reflects website’s responsiveness in providing information in real time conditions.

Table 3: Sub-criterion dimensions of **structure and design**. Dimensions form the sub-criteria identified in Figure 1.

Appearance and Multimedia

Is information displayed in plain, simple and concise text [13]? Are image and icons consistent, e.g., is the same icon used for the same purpose [4]? Have graphics been designed to meet user’s needs, habits and expectations [22]? Appearance and Multimedia captures aspects that relate to site’s “look and feel” with special emphasis in state of the art graphics and multimedia artefacts. It is included in the list of criteria to place emphasis on the fact that multimedia “gadgets” should be carefully used. Dimensions carried through this criterion are emphasized, and often over-emphasized, in relevant literature and relate to the use of icons, colors, text readability from normal viewing distances, viewing by users with special ability characteristics etc. [4, 13, 26]. Dimensions, or sub-criteria, of appearance and multimedia are summarized in Table 4.

Sub-criterion	Definition
<i>Graphics representation</i>	Reflects appearance, usefulness in navigation and contribution of graphics, to the site’s purpose.
<i>Readability of content</i>	Reflects the easiness of reading from normal viewing distances.
<i>Multimedia: image, voice and video data</i>	Reflects the way that combinations of image, voice and video contribute to site’s usefulness and easy of use, and also the way they are appropriately incorporated by the website.

Table 4: Sub-criterion dimensions of **appearance and multimedia**. Dimensions form the sub-criteria identified in Figure 1.

Uniqueness

Uniqueness refers to user’s perception that the site carries something that makes it different in a world full of sites. Site distinctiveness is judged according to content, aesthetics and design characteristics – see Table 5.

Sub-criterion	Definition
<i>Uniqueness of content</i>	Reflects originality of provided information.
<i>Aesthetics of content presentation</i>	Reflects site’s appearance in general overview.
<i>Uniqueness of design characteristics</i>	Reflects the originality of website’s structural characteristics, which should be both unique and aesthetic.

Table 5: Sub-criterion dimensions of **uniqueness**. Dimensions form the sub-criteria of Figure 1.

RATIONALE AND PURPOSE

An initial version of the model criterion space (Figure 1) was critically examined by a small group of web users. During this process, several issues, both with respect to criteria and hierarchical grouping were cleared and redefined. Then criterion, sub-criterion and company website specifics were presented to a homogeneous group of users, drawn from the student population of the Technical University of Crete, Chania, Greece, and composed of 133 fourth semester students from the Departments of Production and Management Engineering and Mineral Resources Engineering. Student selection was done carefully to ensure uniformity in web experience and willingness to participate in experimentation. Initially, users were asked to assess criterion and sub-criterion weight values without knowing that their assessment would later be used to comparatively score the three cellular phone companies.

METHODS

We used AHP to elicit criterion and sub-criterion weights and to assess user preference with respect to the selected company websites. One hundred and thirty three students participated in the process; however, results reported herein are based on 122 student assessments – we discarded eleven assessments because of low consistency ratings in using the AHP. All students were tutored in AHP using the Expert Choice [8] software implementation before using the method to assess criterion weights and selected websites. In addition, students were not allowed to revise assessments. On the average each student spent about 50 minutes assessing criterion and sub-criterion weight values and comparing company websites. Students were asked to use the sites to learn about the services and the alternative programs offered by each company. The process was supervised and students were not permitted to exchange any information among them. Criteria and sub-criteria were presented in different order to students to avoid ordering bias and to explore differentiation in final assessments.

Our work focused on evaluating the following research questions:

1. What is the significance of each criterion or sub-criterion in assessment of website quality?
2. Does criterion and sub-criterion significance differ among users?
3. Can selected criteria and sub-criteria be aggregated to composites that would facilitate website quality assessment in the future?
4. Do selected criteria support rating of the selected company websites?

Analytic Hierarchy Process (AHP)

AHP is based on the idea that a complex issue can be effectively examined if it is decomposed into its parts [9, 25]. AHP entails a hierarchy whose top level reflects the overall objective: the focus. Criteria on which the focus is dependent are listed at intermediate levels, while the lowest level includes the alternatives, e.g., the cellular phone company websites. An element at a higher level is said to be a governing element for those elements at the lower level. Elements at a certain level are compared to each other with reference to their effect on the governing level. Let us consider the elements C_1, C_2, \dots, C_n of some level in a hierarchy and let us denote their normalized, unknown, priority weights by w_1, w_2, \dots, w_n , respectively. The value of w_i reflects the degree of importance of C_i with respect to C_i 's governing element. The first step in the calculation of w_i 's is to derive pair-wise comparisons between the n elements. These pair-wise comparisons are structured into a n -by- n reciprocal matrix $A = [a(i,j)]$, called the judgement matrix. For example, the pair-wise judgements across the five main criteria would be form a 5×5 following judgement matrix in which all diagonal elements would be equal to 1 (this means that $a(i, j) = 1$ when $i = j$) and the valuation of the other matrix elements would be based on the following scale:

$a(i, j) = 1$ if C_i and C_j are of equal importance

$a(i, j) = 3$ if C_i is weakly more important than C_j

$a(i, j) = 5$ if C_i is strongly more important than C_j

$a(i, j) = 7$ if C_i is very strongly more important than C_j

$a(i, j) = 9$ if C_i is absolutely more important than C_j

$a(i, j) = 2,4,6,8$ are used to compromise between two judgements.

In addition, $a(i, j) = \frac{1}{a(j, i)}$. This matrix produces one real valued eigenvector [25], which is used to estimate criterion value weights. Companies are compared in a similar way with respect to each sub-criterion i.e., with respect to “search engines” company A is rated versus companies B and C and so on. The scale of measurement extends from 1 to 9; for instance, a rating of 5 indicates strong preference of

site A versus company B with respect to a sub-criterion. Ratings form a reciprocal matrix and the eigenvector is used for synthesis of results. Synthesis of preferences is achieved by using a linear model, e.g., where S_k denotes the overall score of company k ($k = A, B, C$), w_i denotes criterion (or sub-criterion) weight value and S_{ik} denotes the score of company k with respect to criterion or sub-criterion i .

We do not claim that presentation of the AHP, by means of this section, is complete. Presentation focused in the basics of the method. However, the reader interested in AHP may refer to a wide range of published work, e.g., [9, 25, 33].

Criterion and sub-criterion weight values and company preferences derived from each participant were normalized with respect to overall website quality. Normalized values were averaged over the 122 participants. The process is subject to user bias and error. Saaty [25] has proposed a consistency ratio to steer the process. From a mathematical point of view the “ideal” value for the consistency ratio should be equal to zero; nonetheless values that are less than 0.1 are acceptable. We discarded 11 user judgments because their consistency ratio yielded values greater than the acceptance threshold.

RESULTS

Following user elicitation of judgments on criterion weight values and company preference ratings, data were coded and statistically processed using SPSS. Initial judgment data included 2928 sub-criterion weight values (that is 24 X 122) and 610 criterion weight values (that is 5 X 122). Criterion, sub-criterion pair-wise and overall judgements and eventual company preferences were collected using Expert Choice software [8]. All criterion weight values and preference ratings were normalized with respect to website quality. Normalization of main (e.g., second level) criterion values was done with respect to perceived website quality (the overall objective) while sub-criterion weight values were first normalized with respect to the respective main criterion and then with respect to overall objective. For instance, uniqueness of content, aesthetics in content presentation and uniqueness of design characteristics weight values are normalized with respect to uniqueness (Figure 1 and Table 5).

Dataset	No. of items	Responders	Alpha value
Main criteria level	5	122	0.8920
Sub-criteria level	24	122	0.9112
Company ratings with respect to each sub-criterion	72	122	0.8346

Table 6: Estimation of Cronbach’s alpha values with respect to study datasets. A Cronbach’s alpha value greater than 0.70 means that data are reliable and can be readily used to support analysis [10].

Website preferences yielded a database of 8784 observations (including company ratings with respect to each sub-criterion, that is, 3 X 2928). Evaluation of internal consistency of user judgments revealed three Cronbach’s alpha estimates, which are summarized in Table 6, which suggest that judgments are “tightly connected,” which, in turn, means that experimentation achieved a high degree of response reliability and order of criterion and sub-criterion presentation did not bias user assessment.

Main criterion weight value and company preference statistics are summarized in Table 7. Users rated content as the most significant criterion and placed structure and design as second with navigation coming

to a close third position. Appearance and multimedia and uniqueness and scored low values. Company A outranked company C in all criteria and outperformed company B in all but one criteria, i.e., uniqueness. Company A outperformed companies B and C achieving an overall score of 0.41 – see Table 7.

Across sub-criteria, utility of content achieved top rating with completeness of information, and reliability of content following. Top three sub-criteria are part of the font-runner main criterion, which is content, and account 0.36 of the total weight; they are the only sub-criteria that scored more than 0.10. Sub-criterion weight value statistics along with respective company ratings are summarized in Table 8.

Differences between assessments of criterion weight values, aggregate company ratings and company ratings with respect to each main criterion as well as between criterion and sub-criterion weight values were examined using ANOVA. Results are summarized in Table 9.

Company	Criterion					
	Content	Navigation	Structure	Appearance	Uniqueness	Overall score
A	0.40±0.15	0.39±0.15	0.41±0.14	0.46±0.16	0.35±0.16	0.41±0.13
B	0.36±0.17	0.38±0.17	0.34±0.15	0.31±0.16	0.39±0.17	0.35±0.15
C	0.25±0.14	0.23±0.10	0.25±0.10	0.23±0.11	0.26±0.12	0.24±0.08
Criterion statistics	0.47±0.08	0.18±0.07	0.19±0.09	0.08±0.05	0.09±0.09	$\sum \approx 1.00$

Table 7: Ranking of companies with respect to criteria and with respect to the overall objective (website quality). Entries correspond to mean and standard deviation values. Last column reflects the overall score of each company. Last line reflects mean criterion weight values. Criterion weight values and overall ranking of companies with respect to criteria are normalized to the overall objective. The summation listed in the last line of the table reflects both company rankings and criterion weight values.

Analysis of main effects showed that at the $p = 0.001$ level, effects between criteria and sub-criteria are significant; in addition, company score differences were found significant at the sub-criterion level. Furthermore, results confirmed the hypothesis that variation across criterion or sub-criterion weight values are significant at the $p=0.001$ level. Results summarized in Tables 7, 8, and 9 confirmed criteria and sub-criteria validity. As expected both criteria and sub-criteria proved conducive to giving users the chance to discriminate between website features and between the selected company websites. Additional ANOVA by separating male and female responses indicated about the same patterns of variation. Therefore, it was concluded that there is no need to separate results.

Factor analysis across of company scores, across criteria and sub-criteria revealed two and nine factors explaining 53.4% and 68.9% of total variance, respectively. Results are summarized in Tables 10 and 11. Factor analysis supported the hypothesis that there exist significant commonalities both between criteria and sub-criteria.

DISCUSSION

Factor analysis and loadings over the sub-criteria (Table 11) identified nine factors explaining 68.9% of the total variance. Nineteen of the twenty four sub-criteria achieved significant factor loading. Analysis at

the sub-criterion level increased percentage of variance explained. This result was expected since sub-criteria provide a more detailed account of website quality assessment. In addition, analysis at the sub-criterion level revealed that membership to factor formation was unique; in other words a sub-criterion participated in only one factor

Sub-criterion	Company rating with respect to each sub-criterion			Sub-Criterion weight value statistics
	A	B	C	
Utility of content	0.40±0.13	0.38±0.00	0.20±0.09	0.15±0.07
Completeness of information	0.41±0.13	0.37±0.15	0.24±0.08	0.11±0.07
Subject specialization	0.39±0.14	0.37±0.13	0.26±0.09	0.07±0.05
Reliability of content	0.45±0.13	0.32±0.18	0.23±0.06	0.10±0.07
Syntax of content	0.46±0.14	0.34±0.16	0.22±0.09	0.03±0.03
Convenience of navigation tools	0.48±0.14	0.33±0.16	0.19±0.13	0.05±0.03
Identity of site	0.40±0.14	0.36±0.13	0.24±0.13	0.01±0.01
Means of navigation	0.39±0.13	0.37±0.14	0.22±0.12	0.02±0.01
Links to other sites	0.41±0.13	0.36±0.13	0.20±0.09	0.02±0.01
Ease of use of navigation tools	0.35±0.14	0.37±0.13	0.23±0.13	0.02±0.02
Search engines	0.44±0.13	0.34±0.14	0.23±0.05	0.05±0.03
Order of elements	0.44±0.13	0.36±0.15	0.19±0.09	0.04±0.03
Loading speed	0.46±0.12	0.35±0.14	0.21±0.08	0.04±0.03
Site map	0.46±0.13	0.33±0.19	0.23±0.04	0.03±0.03
Information structure.	0.35±0.15	0.39±0.17	0.27±0.09	0.02±0.02
Software. requirements.	0.35±0.13	0.38±0.16	0.29±0.09	0.02±0.02
Browser compatibility	0.37±0.13	0.34±0.16	0.29±0.04	0.02±0.02
Real time information	0.43±0.14	0.37±0.15	0.23±0.08	0.02±0.02
Graphics representation	0.39±0.14	0.34±0.18	0.26±0.09	0.03±0.03
Readability of content	0.45±0.13	0.30±0.19	0.24±0.08	0.03±0.02
Multimedia: image, voice and video data	0.36±0.13	0.37±0.16	0.26±0.04	0.02±0.02
Uniqueness of content	0.35±0.14	0.36±0.19	0.30±0.08	0.05±0.05
Aesthetics of content presentation	0.36±0.13	0.34±0.16	0.22±0.06	0.02±0.02
Uniqueness of design characteristics	0.39±0.14	0.37±0.19	0.19±0.06	0.02±0.02
Overall company score	0.41±0.11	0.35±0.17	0.24±0.09	$\sum \approx 1.00$

Table 8: Company ratings across evaluation sub-criteria and sub-criterion weight values. Sub-criteria are described in Tables 1 – 5. Entries correspond to mean and standard deviation values. Last column reflects the overall sub-criterion weight value statistics (mean ± standard deviation). Last line reflects overall company score averaged over the 122 students – overall scores are also reported in Table 7. Sub-criterion weight values and overall company scores are normalized with respect to the overall objective. The summation listed in the last line reflects both summation over sub-criterion weight values and company overall scores. Lines separate sub-criteria according to the main reference criterion; for instance, first five sub-criteria correspond to Content (see Table 1), etc.

Source of Variation	SS	df	MS	F
(a)				
Companies	0.3288	2	0.1644	78.197
Main Criteria	4.2506	4	1.0627	505.455
Companies X Main Criteria	0.1982	8	0.0248	11.783
(b)				
Companies	0.0355	2	0,02	3.01
Sub-criteria	1.0782	23	0.05	7.97
Companies X Sub-Criteria	0.3179	46	0.01	1.17
(c)				
Criterion	12.122	4	3.03	492.4
Sub-criterion	3.48	23	0.15	124

Table 9: Results of two way ANOVA. Part (a) of the Table refers to ANOVA (with replication) between companies and main criteria. Part (b) of the table refers to ANOVA (with replication) between companies and sub-criteria. ANOVA was performed on company scores normalized with respect to the website quality. Effects between criteria and sub-criteria are significant; in addition, effects in company scores are significant at the sub-criterion level. Part (c) of the table presents ANOVA results (without replication) on criterion and sub-criterion values – sub-criterion weight values were normalized with respect to the website quality. All effects were studied at $p = 0.001$.

Criteria	Factor		
	1	2	h ²
Content	0.72	0.42	0.69
Navigation	0.70	0.20	0.53
Structure and design	0.04	0.79	0.62
Appearance and multimedia	0.16	0.56	0.34
Uniqueness	0.64	-0.42	0.59
% of Variance	33.65	21.71	
Total Variance			55.4

Table 10. Rotated factor loadings obtained for the five criteria; h² values in the last column are communality values. This measure indicates the strength of the relationship between rotated factor loadings and criteria. Values less than 0.5 are considered signals of a weak relationship between a criterion and the factor upon which it loads. Loadings with respect to each factor – criterion pair are boldfaced to highlight membership.

Rotated factor analysis across sub-criterion weight values indicates that:

- Factor 1 contains sub-criteria belonging to content and navigation; the specific criteria achieved a total of about 18% in total weight. Furthermore, analysis at the criterion level grouped content and navigation together into Factor 1 formation.
- Factor 2 contains sub-criteria drawn from navigation and structure and design. At the criterion level navigation and structure and design are not part of the same factor (Table 10); however, grouping them into a common factor at the sub-criterion level may be seen as contributing to the increase in percentage of variance explanation. Indeed, the two sub-criteria account a total of about 7% in sub-criterion weight assessment.

- Factor 3 includes the three sub-criteria originally identified as part of uniqueness. The three sub-criteria account a total of about 9% in total weight assessment.
- Factor 4 includes sub-criteria that are part of content and appearance and multimedia, which in part explains the loading of the latter in factor 1 formation at the criterion level (Table 10). Total weight carried by these two sub-criteria is about 13%.
- Factor 5 includes sub-criteria that were originally part of structure and design alone – this result enhances criterion factor formation while total weight carried is about 5% .
- Factor 6 includes sub-criteria that were originally part of structure and design alone – this result, similarly with factor 5 formation, enhances criterion factor formation while total weight carried is about 6%.
- Factor 7 combines two sub-criteria, which were originally part of content and structure and design. Given that at criterion level the two form into different factors we attribute this formation, similarly to factor 2 formation, to the increase in percentage of variance explained at the sub-criterion level of analysis. Total weight carried by this factor is equal to about 9%.
- Factor 8 includes one sub-criterion from appearance and multimedia, it accounts to about 3% of total weight value and explains partly the non significant loading of the respective criterion to any factor during analysis at the criterion level.
- Factor 9 includes one sub-criterion from content; however, this is the top runner sub-criterion (Table 8) and carries about 15% of total weight value.

CONCLUSIONS

Analysis of results demonstrated significant synergy among criteria and sub-criteria, which confirms our initial hypothesis about criterion interdependency. Criteria grouped into two factors and sub-criteria analysis led to the formation of nine factors. Factor analysis identifies nine generic criteria, which capture the all significant aspects of website quality assessment. Each of these criteria carries an indicative weight value; however, further research would be necessary to validate composite criterion weight values. A composite criterion weight value is calculated by summing up over the sub-criterion weight values that are part of the new criterion. For example, the four sub-criteria are part of factor #1 (Table 1) have corresponding weight values that sum-up to 0.18 (see Table 8).

The nine composite criteria that correspond to the result of factor analysis summarize research findings, namely (in parenthesis are listed the weights taken from Table 8):

1. **Relevance (18%).** It relates to the perception the website creates to the visitor regarding the significance of its content to the visitor's inquiry. For instance, if the visitor is looking for music then a high degree of relevance is achieved if the website contains information about the type of music the visitor is interested in. Relevance represents a rather objective dimension. The website is relevant according to the degree that the information it contains is relevant to visitor's interests.
2. **Usefulness (15%).** Usefulness extends relevance to the nature of the specific visitor's inquiry. For instance, the site may contain information about the music the visitor is interested in; however, may or may not contain information, which is practical to visitor's inquiry and needs. To this end, usefulness emerges as a rather subjective dimension; however, website developers should continuously check information contained in their site to assess usefulness to a wide audience of visitors. Often we see website administrators asking visitors to assess information while they are visiting a site, or, providing star based qualification of site material.

3. **Reliability (13%).** Reliability relates to accuracy of information contained in the website. The dimension is rather objective. For instance, a site focusing on providing information about music of a certain type should also guarantee accuracy of content. Often designers include a note about last update of information. Doing so they help the visitor in forming an opinion regarding website reliability.
4. **Specialization (9%).** Specialization captures the specificity of information contained in the website. In the analysis of results reported in this article specialization achieved high importance. It means that information should incorporate all necessary details that visitors may be looking for. For instance, a website incorporating music information should include date and place of recording, equipment used, etc. Specialization contributes to site relevance and usefulness yet places a heavier burden on reliability.
5. **Architecture (9%).** This dimension captures the organization of objects via which information is conveyed to the visitor. Placement of buttons, colors, special effects and the like are part of architecture. It came to no surprise to us that analysis of experimental results did not yield a higher importance to architecture. Almost all sites floating around make use of advanced graphics, images and often video and sound. Thus visitors are no longer so much impressed by these characteristics.
6. **Navigability (7%).** This dimension reflects both the easiness and convenience of moving in and around the site.
7. **Efficiency (6%).** This dimension captures the technical performance characteristics of the website; is it slow, is it fast, does the visitor get advance notice about the estimated time it may take to retrieve information, etc.
8. **Layout (5%).** This dimension reflects on the unique aspects involved in website objects presentation. It relates to architecture yet it is used to differentiate the site for the really unique design characteristics it possesses. Yet the rather low significance it achieved during experimentation proves that visitors are no longer impressed with “bells and whistles.”
9. **Animation (3%).** A rather insignificant dimension that captures the moving aspects involved in presentation of information and website – user interaction.

The work and results reported herein were based on a uniform collection of websites and a rather homogeneous group of users. We did so to limit results from effects from a variety of uncontrollable factors. We started from a large set of criteria and sub-criteria and derived a comprehensive set of composite criteria. The process via which judgement about the websites was elicited was based on the calculation of trade-offs between criteria, sub-criteria and preference about websites. The work resulted in the formulation of a quality assessment model, which is based on nine composite criteria. These criteria focus more on the semantics that underlie website use and address to a lesser degree the software engineering characteristics of the site.

The criterion proposal identifies two areas for further research on the subject. The first area follows directly and relates to the formation of a balanced website goodness model using the nine criteria. Such an endeavor will also contribute to the validation of the composite criterion weight values and lead to the establishment of objective benchmarking across websites. The process will also link website assessment to customer satisfaction and customer relation management. The second area for further research on the subject links to the verification and validation of the nine composite criteria using other types of websites and specifically websites, which are enhancing site – user interaction and information exchange. Results reported herein were largely based on service topical category sites targeted to inform the visitor rather than to sites oriented towards collecting and processing visitor information.

Specific factor loadings with respect to criteria are reported in Table 10. Factor 1 captures content, navigation and uniqueness. Factor 2 includes structure and design – appearance and multimedia loaded more on factor 2 than to factor 1. Users perceived content and uniqueness as tightly connected and

associated both to navigation giving a gestalt view of website quality that captures the “what” material the site offers and “how” the user may navigate around the material. Factor 2 encompasses the design of the website, e.g., how the site looks like to the user.

Factors	Sub-criterion factor loadings	Variance explained (%)
1	Completeness of information (0.68) Syntax of content (0.73) Means of navigation (0.77) Links to other sites (0.66)	11.66
2	Convenience of navigation tools (0.71) Software requirements (0.88)	9.27
3	Uniqueness of content (0.85) Aesthetics of content presentation (0.76) Uniqueness of design characteristics (0.69)	8.21
4	Reliability of content (0.70) Readability of content(0.77)	7.63
5	Site map (0.86) Information structure (0.85)	7.60
6	Loading speed (0.74) Real-time information (0.74)	6.82
7	Subject specialization (0.77) Browser compatibility (0.70)	6.66
8	Graphics representation (0.86)	5.82
9	Utility of content (0.67)	5.19
	Total variance explained by all factors	68.9%

Table 11. Results obtained from rotated factor over sub-criterion weight values analysis. Presentation is limited to sub-criteria, which achieved a communality value (h^2) greater or equal to 0.5 – communality values are listed in parenthesis next to each sub-criterion.

BIBLIOGRAPHY

- [1]. Athanasou, J. “A framework for evaluating the effectiveness of technology assisted learning.” *Virtual University Journal*, 2. 1999. pp. 13-21.
- [2]. Bauer, C. and Schari, A. “Quantitative evaluation of web site content and structure.” *Internet research*, 10, 2000. pp. 31-44.
- [3]. Beck, S. "Evaluation Criteria: The Good, the Bad and the Ugly: or, why It’s Good Idea to Evaluate Web Sources.” At website: <http://lib.nmsu.edu/instruction/evalcrit.html> (visited in August 2003).
- [4]. Borges, J., Morales, I., and Rodriguez, N. *Page design guidelines developed through usability testing*. Lawrence Erlbaum Associates, New Jersey. 1998. pp. 137-152
- [5]. Bramley, P. “Evaluating effective management learning.” *Journal of European Industrial Training*, 23 (3). 1999. pp.145–153.
- [6]. Cosmote. *Official Website of Cosmote*. Website: www.cosmote.gr (visited in August 2003)
- [7]. Derek, A. and Tanniru, R. *Analysis of customer satisfaction data*. American Society for Quality Press. Milwaukee, USA. 2000.
- [8]. Expert Choice. “Official website of the provider of the Expert Choice software.” Website: <http://www.expertchoice.com/> (visited in August 2003).
- [9]. Forman, H. “Decision support for executive decision makers information strategy.” *The Executive’s Journal*, 1985. pp. 4-14.
- [10]. Frankfort-Nachmias, C., and Nachmias, D. *Research methods in social sciences* (4th ed.). St. Martin’s Press Inc. London, 1992.

- [11]. Gattorna, L., and Walters, W. *Managing the Supply Chain: a strategic perspective*. Macmillan Business. Great Britain, 1996.
- [12]. Gauch, S., and Xand, Z. "Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web." In: *Proceedings of the 23rd Annual International ACM/SIGIR Conference* (pp. 288–295). Athens, Greece, 2000.
- [13]. Grose, E., Forsythe, C., and Ratner, J. (1998). *Using Web and traditional style guides to design web interfaces. Human Factors and Web Development*. Lawrence Erlbaum Associates. New Jersey, 1998. pp.121-131.
- [14]. Hall, H. "Imaging and textual components of web page design." *Virtual university journal*, 2. 1999. pp.58-62.
- [15]. Herczeg, M., and Kritzenberger, H. (2001). "A task and scenario based analysis model for user-centred systems." In: *Proceedings of the 9th international conference on Human-Computer Interaction*. New Orleans, 2001. pp.229-235.
- [16]. International Organisation for Standardisation, ISO *ISO/IEC FCD 9126-1.2. Information Technology - Software Product Quality. Part 1: Quality Model: Draft*. 1998.
- [17]. International Organisation for Standardisation, ISO *ISO/IEC 14598-3. Information Technology - Software Product Evaluation - Part 3: Process for Developers. Software Engineering: Draft*. 1998.
- [18]. Ivory, M., Sinha, R., and Hearst, M. "Empirically Validated Web Page Design Metrics." In: *Proceedings of the ACM SIGCHI'01 Conference: Human Factors in Computing Systems*, 5360, New York, 2001.
- [19]. Kanerva, A., Keeker, K., Ridsen, K., and Schuh, E. (1998). "Web usability research at Microsoft Corporation". In: Grose, E, Forsythe, C., & Ratner, J. (Eds), *Human Factors and Web Development*. New Jersey: Lawrence Erlbaum Associates. New Jersey, 1998. pp.189-198.
- [20]. Korpela, J., and Lehmusvaara, A. "A customer-oriented approach to warehouse network evaluation and design." *International journal of Production Ergonomics*, 59, 1999. pp.135-146.
- [21]. Losavio, F. "Quality Models to Design Software Architecture." *Journal of Object Technology*, 1 (4), 2002. pp. 165-178.
- [22]. Mitrovic, N., and Mena, E. "Improving User Interface Usability Using Mobile Agents." In *Proceedings of the 10th DSV-IS Workshop*. Madeira Island, Portugal, 2003.
- [23]. Nielsen, J. *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing. Indianapolis, USA., 2002.
- [24]. Ratner, J., Grose, E., and Forsythe, C. "Characterization and assessment of HTML style guides." In *Proceedings of ACM CHI 96 Conference on Human Factors in Computing Systems*, (volume 2), 1996. pp 115–116).
- [25]. Saaty, L. "Axiomatic foundation of the Analytic Hierarchy Process." *Management Science*, 32 (7), 1986. pp.841-855.
- [26]. Shneiderman, B., *Designing the User Interface: Strategies for Effective Human-Computer Interaction: 2nd edition*. Addison-Wesley Publishers Co., Reading, MA, 1992.
- [27]. Telestet. *Official website of Telestet*. Webiste: www.telestet.gr (visited in August 2003). Currently available at <http://www.tim.com.gr/>
- [28]. The International Academy of Digital Arts and Sciences. "The webby awards 2003 judging criteria." Website: <http://www.webbyawards.com/judging/criteria.html> (visited in December 2003).
- [29]. Virpi, R., & Kaikkonen, A. "Acceptable download times in the mobile internet." In *Proceedings of the 10th International conference on Human – Computer interaction*. Lawrence Erlbaum Associates. New Jersey, 2003. Volume 4, pp.1467-1472.
- [30]. Vodafone. Official website of Vodafone. Website: www.vodafone.gr (visited in August 2003)
- [31]. Vora, P. "Human factors methodology for designing web sites." In Grose, E, Forsythe, C., and Ratner, J. (Eds), *Human Factors and Web Development*. Lawrence Erlbaum Associates. NJ, 1998. pp.189-198.
- [32]. Webber, A., Apostolou, B., and Hassell, M. "The sensitivity of the analytic hierarchy process to alternative scale and cue presentations." *European Journal of Operational Research*, 96. 1996. pp 351-362.
- [33]. Wind, Y., and Saaty, L. "Marketing applications of the analytic hierarchy process." *Management Science*, 26 (7). 1980. pp. 641-658.