

3. METHODOLOGY

3.1 Establishing Conceptual Framework

Generally this paper has follow the 4 steps Web automatic assessment formula by Brinks and Hofer (2002) begins with deciding on the approaches and the requisites; development of an easy to use and practical evaluation instrument, followed by selection of the domains and lastly configuring the technique framework.

As agreed by many other researchers (von Dran et al., 1999 Gerhke & Turban, 1999, Aladwani, 2002) owing to the fact there is lack of comprehensive method and compromised set of metrics are the intricate impediments. This paper is to execute a two steps approach to close up possible analysis gap by capturing both objective and subjective measure on same set of sample airlines Web sites.

Decision on evaluation instrument:

The theoretical instigating point for this study is inspired by the notion of a two dimensional measure experimented by Rachman and Buchanan (1999) using the Effective Web Model that drill down to the eventual strengths and weaknesses of tourism Web sites by looking at the features design and overall users value perceives, and ultimately reach to a four quadrant Importance-Performance Comparative Analysis.

Rachman's predefined checklist comprises 59 prominent features sourced from the world's best tourism Web sites, has in fact created a benchmark Web design for the tourism industry. Although there are other models emphasized on Web features designs (Gerhke

& Turban, 1999; von Dran et al., 1999 and Kuo et al., 2004), this is the closest to airlines industry and most compromised by the industry experts, users and academicians. Performing features review is very important for benchmarking purposes, that is to keep the 'best practices' of other organizations are integrated into owns storefront too.

Nonetheless, earlier Rachman & Buchanan's fundamental was only assessing the Web design, they have extend the analysis to assess the sites from another angle through user's perceptions. Resemble to their technique, this study is also applying the second phase evaluation using a instrument with users perception conceptualized by Merwe and Bekker (2003) anchored in a marketing theory - the *customer buying process*.

Both instruments are essentially reinforcing each' integrity for achieving the effectiveness objective. Uncovering the variations and connections between two produce results that may highlight properties for industry-specific benchmarks.

One significant advantage of this dual core measurement is that it shows strong element of functional benchmarking and comparative analysis. Each evaluator would have sufficient points of reference.

Why dual evaluation – they are link closely

Well designed Web sites are critical success factor (Kim et al, 2003) for an e strategy realize these benefits, when potential customers access the Web sites, the appearance, structure and maintenance status will influence the user's perception of both transaction experience and corporate image. A result from Forrester Research shows that when consumers cant find an item on Web site, approximately 50 per cent will leave, and if they stumble upon negative

experience on the first visit, 40 per cent will never come back. These are the most worrisome possible scenarios imposed by a poorly designed Web site (Kim et al, 2003)

I expected interface design, navigation and system technicality will be highly correlated with Web features result, suggesting there is one influential underlying them all.

Decision on evaluation criteria:

Building instrument is not any easy task too, it needs be a tool that not too dependent on historical data, flexible enough to accommodate within the fast pace, diffuse and ubiquitous effects of technology advancement. (Kardaun, 1999).

The metrics used in this study included both absolute and relative measures. Web features review represents the absolute metrics whereas relative measures are review from the users while performing task. For example, the ease of finding information on an airline Web site may be perceived differently.

Although Rachman & Buchanan's checklist was benchmark to tourism Web sites, it was however developed some times ago. With technology advances and higher quality demand from users, refinement was done to include some new and relevant features suggested by other papers yet tested reliable. The Web as a whole is to serve customers' need must have profound features for the entire designing principle. New features are added to the shortcoming of having just the 'expected' and 'normal' features. Presence of 'exciting' features may generate user loyalty and are most desirable by businesses to create competitive advantage. (von Dran et al, 1999).

Total of 65 features for examinations are finalized and presented in Appendix 3. Each feature found will be given 1 score.

Here, I made a very important assumption, the decision on Web designs are dependent on the management concerns. Although created by Web master, the underlying strategy is exclusively determined in the hand of management team.

Throughout the selection of subjective measures, there are quite a number of research references attempt to adopt users perception and satisfaction as a subject of interpreting quality of e-commerce Web sites as shown in the early summary table of literature reviews (refer Appendix 1).

In order to have the most relevant and validated criteria of discussion, this paper only take up the fist 5 most frequently used. Although there are criterias share the same frequency, for benefit of fairness, all 19 criterias are absorbed:

- Navigation
- Contents characteristics
- Appearance
- System reliability
- Contents structure
- Usefulness
- Security
- Ease of use
- Help/Support
- Functions
- Content Currency
- Usability
- Interface design
- Purchase agreement
- Purchase settlement

Privacy
Consistency
Page loading
Web interactivity are apply

But, pilot test shows very low Cronbach's alpha coefficient for these criterias. Thereof to recognize a more streamlined pattern, these areas of discussions are regrouped into just 5 constructs namely Interface design, navigation, content characteristics, functional application and the system technical platform (Figure 1). These areas practically meets all the evaluation criteria outlined in earlier section and show higher Cronbach' alpha coefficient than before (Figure 2).

Most Frequent Discussed Criteria

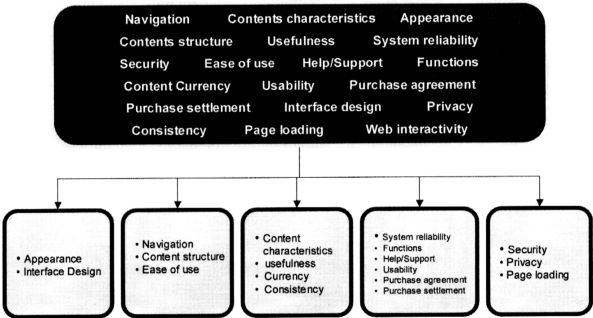


Figure 3 : Regroup of Evaluation Criteria

Justification of choosing Merwe & Bekker's model comes in two reasons. Firstly, airline Web sites are genuinely known as a storefront that require specialty in selling, and this model clarified the buying

process and try to answer how well are the Web site in matching requirement throughout the buying cycle. Second reason, is solely coincidental that all the regrouped criteria are utterly found in this model. However, some of the attributes are removed, specifically from system technicality perspectives. For example, limitation for normal users to determine the technical specs like 'good integration of different systems onsite', and data transfer rate and database sizing.

Criteria	Attribute	Avg α
Interface Design	1-15	0.637
Navigation	16-30	0.627
Content Quality	31-46	0.903
Reliability	47-63	0.851
System Technicality	63-71	0.572
Overall	1-71	0.932

Figure 4 : Summary Of Reliability Analysis For Questionnaire Categories

In fact, comparable criteria were used in industry practices. For example, Webby Award (presented by The International Academy of Digital Arts and Sciences), the leading international award honoring excellence in Web presence has also assess Web sites based on six criteria: content, structure and navigation, visual design, functionality, interactivity, and overall experience. (Webby Award Academy is a US 500-member body of leading Web experts, business figures, luminaries, visionaries and creative celebrities)

The five identified constructs are evaluated by evaluators using questionnaire with 71 attribute. There are 71 questions in total rated on a Staple scale range from -2: very disagree to +2: very agree. Because of the key criteria are selected from the most frequent mentioned and highly recommended constructs for quality assessment, thereof '-2' is be assigned on any unfound attributes because it brings reverse effect

to the aggregate value perceived. Appendix 4 shows a sample questionnaire. The staple scale asks a person to rate a brand, product, or service according to a certain characteristic on a scale from +2 to -2, indicating how well the characteristic describes the Web environment.

3.2 Identification of Sample Web sites

The data collection was designed to capture the broadest random sample of airline Web sites possible to acquire data in various geographical region and segments between low frills and conventional full service airliners (here I call traditional airlines). In such, a comprehensive airlines directory listing (sampling frame) need to be drawn out. Yahoo!Travel Search was able to give a longer and detailed listing compared to other search engine. From the seven continents, search result shows total of 288 airlines are classified. Because of the Internet usage rate and number of airline companies differ across regions, this paper uses proportional stratified sampling method to ensure each regional characteristics will be captured as fair means of comparison. Figure 3 depicts how the proportional stratified method is operated:

Region	Total Airline	Overall %	Sample size
Middle East	20	7%	2
Latin/South America	24	8%	4
North America	51	18%	6
Asia	58	20%	6
Europe	108	38%	8
South Pacific	13	5%	2
Africa	14	5%	2
	288		30

Figure 5 : Sample Sizing

Beside grouped them geographically, airline Web sites are also differentiated by business model – traditional airlines and low cost airlines. Purpose of doing so, is to investigate is there strategy disparity for Web presence.

Determining the sample sizing intuitively is one of this study's weak spot. Although statistically larger sample size would provide higher confidence level of the data accuracy, however in this context there is constraint in both time and resources to achieve.

Among the 30 sample Web sites, number of airlines for each region and numbers of airline for each business model (low cost vs traditional carriers) are proportionately selected. Appendix 5 has listed them in proper categories.

Although the sample units are listed, it is necessary to test and verify if it's currently active and operational before proceeding to data collection. One of the following search engines was used to find the Web page (Misic & Johnson, 1999):

1. Google
2. Excite
3. Netscape
4. Yahoo
5. MSN
6. Mozilla Firefox

The full name of the airline company was entered as search keyword. If the target site appears on top of the 5 matches, it is consider easy to find. An exclusion is made for the airlines type that it must be conventional fleet and not private skyjets companies that also offering

Wfeb sites to their customers. As pilot test, each found page that is active are screened to ensure requirement is met.

3.3 Data Collection

This evaluation platform addressed two stages of data collection. The first fieldwork involves browsing the sample Web sites. Base on the checklist developed earlier, each features found is tick under the 'Yes' checkbox. Each sub page is visited and to reduce the effect from Internet service provider bandwidth congestions, all observation are done almost same time 10am to 11am each day.

While spotting the features is simple and straightforward, the second data collection technique is more task oriented. The outlined evaluation criteria are assessed in sub-element, which I named as attributes. Each evaluator is given 5 Web sites randomly. The measurement questionnaires are distributed with clear indication to what extent they agree or disagree with the descriptions related to the Web site they visited. The composite scores for 71 attributes would differentiate the values perceived.

Due to resources constraints, I only invited 6 voluntary evaluators in reviewing their assessment on the Web sites. Typically each inspection session takes about an hour. However, depending on the complexity and number of dialogue elements, some may consume longer time. In comparison with traditional user testing, having evaluators with prerequisite online experience is able to avoid unnecessary users' reluctance to feedback, typically when they know the objective is to discover mistakes user make while using the interface (Nielsen, 1994).

To avoid complication, assumption is made that user profiles are explicit in resulting variability for the findings. Components for capturing and processing the empirical data are described in the following sub-sections.

3.4 Data Processing

My analysis first begins with using the data collected from reviewing Web page features. A checklist contain 65 features as listed in Appendix 3, were matched against 30 Web sites. Specifically, I need to find out if there are significant variation within single, inter-company profiles; between low cost and traditional airlines; regional comparison and the generalize a statement for the entire industry's Web features design.

Subsequently, data collected from criteria evaluation questionnaires are examined to identify which criteria are most outperform. Data collected are tabulated with SPSS software for descriptive analysis

Third part would be the most crucial analysis to this study. Consolidated data from both model will be compared in percentage basis, reason is because the number of attributes and score assignments are in different scale. Only percentage can do fair comparative analyses. Last part shall end with proposing an interpretation for the meaning of effective Web site.