2   LITERATURE REVIEWS

Question on Web site quality is no longer an infancy research topics, from both academic and industry practitioner, there are loads of studies and approaches have been proposed for Web site evaluations. Take a quick glance, there are models developed for technical Web designing purposes, models to explain Web user and owner behavior, and also Web site performances evaluated from multi dimensional perspectives. Some are domain dominated, some user centric and others business issues related. Basic Web metrics are generally accepted as the start of the Web evaluations concept. Although there are no industry standard definitions for these measures, many had proposed series of criterias supported with empirical data. Bottom line is each aim to know the success and failure of this important communication medium. Nevertheless, success of Web presence is subjective and evocative. It could mean return on investment, profitability, effectiveness, reliability, utility, or competitive advantage. An important point to note is that the characteristics that define success for one Web site do not necessarily define success for another (Phippen, Sheppard & Furnell, 2004). This paper shall prove that careful selection of evaluation metrics and instruments make a difference in distinguish the effectiveness of the carriers web sites.

Generally there are three recurrent themes in Web site evaluation:
    The evaluation instruments
    The evaluation criteria
    Benchmarking

2.1   The Evaluation Instruments

The taxonomies of evaluation tools imply great deal of variations and much dependent on the scholar's concentration, for example scholars from computer engineering background would use automated tools in their investigation whereas students of consumer behavior would skew towards user review technique. But each of these models has its strength and weakness:

1. Automated software

In order to overcome the methodological limitations of subjective impressions, anecdotal evidence, or convenience samples, a number of automated systems have been introduced as well. The research basis is to use framework and a software prototype imply the definition of measurable and operational criteria. These criteria will be investigated and preprocessed by the tool for each Web site separately. This model parameter is technically expressed in coding formulas. In this context, Bauer & Scharl (2000), Brinck and Hofer (2002), Phippen, Sheppard and Furnell, (2004) for example, suggested three domain independent criteria, which are content, interactivity and navigation to support automated software for classifying and evaluating quantitatively wide array of Web sites. While Olsina and Rossi (1999) developed a Web-based tool to supports the administration of evaluation called Web Quality Evaluation Method (WebQEM) to detect either absent features, or poorly implemented requirements. Ivory, Sinha & Hearst (2001) also established an automated tool called Tango (Tool for Assessing Navigation and Information Organization) for usability evaluation.

This notion is also referenced as Web analytics and commonly practiced by many large corporations using custom key performance indicators like log files analysis, click stream analysis, traffic base analysis (e.g. page tagging, page hit) and error reported that are related to present prescriptions of the Web site's infrastructure. Drawback of this approach is the difficulty of obtaining other company's internal data for comparison. Although the results is useful for self-improvement, but it is less effective as tool for securing competitive advantage.

2. *Human involvement*

The second notion's center of attention is the dependence on subjective assessments in the form of user judgments. Relatively subjective instrument that allow users interpret the meaning of quality through their perceptions. Why user is important? Creating a satisfying online experience in a Web site is one of the many goals Web designers. Satisfied users may spend longer time and revisit the Web site continuously, and may even recommend to others (Zhang and von Dran, 2000). Therefore, it is useful to determine what makes a user satisfied with a Website as well as what may potentially dissatisfy. And it was proven successful Web site has high retention rate by securing customer satisfaction (Hung & McQueen,2004; Barnes & Vidgen,2002).

The following illustration is depicted from the renowned consumer research agency Boston Consulting has develop a virtuous circle of satisfying online customers. The viral connections explained why interruption in any point is reference

to "Continuous Improvement to online experience", and it speaks for the survival of entire online business activity.

Source: The Boston Consulting Group, 2001
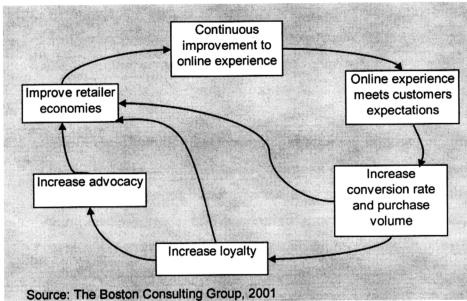
Figure 1 : Virtuous circle of satisfying online customers

Anyone can develop a web site that loads and runs reasonably well, and might even look attractive to the untrained eye. But the design and development of an effective web site requires an assessment of the audience (Nielsen, 1992), an identification of the common tasks that they will perform, and an empirical evaluation of the web site's usability to assess if usability goals have been met. The motivation, reading level, color sensitivity, culture, primary language, software knowledge, and background experiences of users is important as it alters the solution set for the design. (Calongn, 2001)

Within the human computer interaction context, user assessment is necessary in order to predict satisfaction and acceptance of the web environment. The common uses of instruments are online survey, heuristics evaluator, Delphi panel

and features inspections. Closer to recent, there are more empirical investigations examining users satisfaction from different aspect of a web site. (Ho,1997; Tang & Bell,1998; Misic & Johnson,1999; Frokjer, Hertzum and Hornbaek,2000; Srivihok,2000; Kim et al,2003; Zhang & von Dran,2000; Hung & McQueen,2004). Unfortunately, many have just managed to disclose the weak spot and not be able to suggest practical solutions.

On top of just human computer interaction, from marketing and consumer behaviors point of view also developed numerous theories and framework explaining the online purchasing pattern and marketing movements. The broader evaluation criteria need s to implicitly recognize delivery of service value are different across differ customers.

Users, beside important to comment on the performance of certain Web site, they also play crucial role in validating constructs developed for web design guidelines and evaluation criteria for pre-release software. Sample studies include also automated software development like found in EWAM model by Schubert & Selz(2000), 2QCV3Q meta model by Mich, Franch, Novi & Marzani (2003) and Phippen et al (2004)

While user's opinion is important, selection of user are also important, Kardaun has proposed a consumer panel approach by monitoring sample of population's internet traffic. However, this type of longitudinal research is too costly and lack of spontaneous response from users.

Perhaps the main reason why basic metrics are now considered inadequate is the increased involvement of the

customer in a Web experience. It is difficult and expensive to attract customers, but easy to lose them. The Internet has facilitated this – there has been a customer revolution. Customers are no longer confined to their locality in the search for goods and services: they can easily browse nationally, and even internationally. Customers have been empowered – they can choose and browse via the Internet. In addition, any business conducted over the Internet now involves the customer to a much greater degree than ever before. (Phippen et al.,2004)

However, peers studies using objective statistical measurements and subjective form of individual preferences becomes less effective in valuating Web sites due to incremental complexity of the sites and individual preferences may not applied to contemporary Web Commerce considering there are more business related criteria are required now (Hung & McQueen,2004; Kim et al.,2003) A more robust instrument for technology advancement and ever evolving web environment would requires an integrative analysis from the human and automated web analytics.

2.2 The Evaluation Criteria

The second theme most discussed among researchers is regards to evaluation metrics – what should be the assessment parameter in order to accurately produce the meaning of success in each own definition.

A snapshot of the metrics used throughout the discussed literatures is appended in Appendix 2. Apparent statistics shows that 'navigation', 'content characteristics' and ' 'appearance' are the first 3 most discussed criteria. In contrast, 'credibility', 'ability to handle exceptions' and 'visual effect' is the least mentioned among 22 literatures went through. Having said so, reliability of this chart is still relatively insufficient to draw a conclusion, naming convention could be one of the biggest barriers in drawing the line between, good example are 'visual' and 'appearance', 'credibility' and 'reliability', all these names are somehow related and confuse readers without any strong technical knowledge.

Despite the less organized summary, there are still studies share certain degree of similarities in each conceptual model. The common ones are no stranger but closely related to usability and designs (Hung and McQueen, 2004).

Metrics used for evaluating designs are rather straightforward. In this part of investigation, the underlying assumption is that quality of a web site is possibly recognized through the designs and available features. In this aspect specifically, von Dran et al, (1999); Keevil, (1998); Kuo, Hwang & Wang (2004), tried to evaluate a Web sites using interface design checklist. Interestingly is the discussion by von Dran and Zhang (2000) on factors that affect user's online satisfaction and dissatisfaction. The basis of this research was primarily to identify which are the missing features that deteriorate functional and

serviceable manner (hygiene factors), and which are the features that add values and enhance satisfaction (motivator).

Resemble to technology, approaches to evaluate web designs evolve over time. At early stage, studies were concentrating on individual page metrics (Sinha et al, 2001). While other paper suggested to shift the focus from individual pages to aggregated collections based on Web directories, domains and entire site. (Thelwall, 2003). Some even stripped down the investigation by studying the level of quality Web features that satisfies basic, normal and excitement needs of potential users (Zhang & Small, 1999).

While designs principle is accentuated, Barnes and Vidgen (2002) has different stance that usability is about how user perceive and interacts with a Web site, how users experience the navigations is most important rather than design quality principles that funnel the Web designers. Nielsen (1999), Gerhke & Turban (1999), Mitch et al (2003) on the other hand had recommended a fruitful range of usability principles. Usability has always been on the top of the chart for web evaluation, Selvidge (1999) too examined the usability of seven airline Web sites on the basic task of locating flight and fare information.

Others focuses were also drawn base on Web functionality and transactional phases like Schubert and Selz (1999), Kim et al.(2003), Merwe and Bekker (2003), Piccoli, Brohman, Watson and Parasuraman (2002), Hung and McQueen (2004). The concept derived from the fact that Web site is the important platform linking up the communications between business and customers and for airlines industry specifically in this study is generally categorized as a transactional e commerce model and agreed as to sell seats (Phippen et al.,2004) Therefore information collected from Web sites will be the baseline to improve effectiveness of transactional phases. Figure 2 is extract from Hung & McQueen's theory how the criteria in measuring

delivery value depends on degree of satisfaction user perceived when tumbles upon failure within the 3 points. In general, the smoothness of accessing the site and information, effectiveness of feature functionality and navigations will determine user satisfaction then representing Web quality.
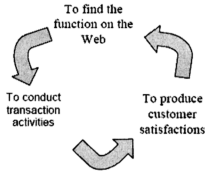
Figure 2 : Evaluation criteria –interactions between Web functions, transactional phases and customer satisfaction

.   Criteria also play important roles for automated software development in defining the parameters, Kano model for example had adopted competitive analysis, scenarios inspection, log analysis, A company that operates mainly informational Web site could perceive the site as successful if it attracts large numbers of people. However, for companies using their Web site as a transactional success may be better defined by a large number of purchases being made from it. There needs to be the means for an organization to measure how successful their Web strategies are if they are to be able to assess the satisfaction of their aims effectively. Generally, at present, this measurement comes from basic, easily obtained statistics such as hits, page views, visits, visitors and sessions, etc. (Phippen, Sheppard & Furnell, 2004).

In company to the criteria discussed, focus on the Web sites effectiveness is also a well-liked topic. Kim et al (2003), Bell and Tang (1998), Rachman and Buchanan (1999) examined the effectiveness of Web sites base on criterias namely the business functions, corporation credibility, content reliability, web site attractiveness, structures and navigation, predominantly from the user's perspective. While majority conceptualize users as potential customers, Hahn and Kauffman (2003) presented differently that users are considered as the co-producers. It is from the theory in production economics to offer a basis for measuring the effectiveness of e-commerce Web sites. Since the customers' efficiency and productivity also become important precursors to high quality service offered by the Web sites.

2.3   Benchmarking

Benchmarking has found its way through contemporary strategic planning, to discover the 'best practices' of other organizations and find way to integrate these qualities into owns operations. Retrospectively, benchmarking has tapped on many areas of an organization, including information systems, unfortunately rare effort is available specifically to address the growing elements World Wide Web. Present Web sites are much more than just graphically-oriented hypertext but the cheapest most cost effective storefront and up to the organization to decide how sellable is their "electronic" sales agents. It is critical for management to know relative positioning from others and what new functionality or features necessary to maintain an edge (Misic and Johnson, 1999; Barnes & Vidgen, 2002).

By and large, some guidelines are treated as benchmarks too. These type of discussions mainly furnished with theoretical guidelines without the support of empirical data: (Calongn, 2001, Brinck & Hofer 2002, Bevan,1995) Ultimate findings is not related to how well or poor Web sites are doing, but more towards understanding what are the quality aspects that users think important and not in evaluation process. The shortcoming of this type of benchmarking is too domain dependent, too general in applying it for specific industry sites.

Slightly different from the guideline-base, another type of benchmarking effort is made through a collection of criteria with an index for users to weigh the importance in the evaluation model. The WAM (Web Assessment Model) developed by Gerhke & Turban (1999) and WebQuol by Barnes and Vidgen (2002), Schubert and Selz (1999), had contributed significantly in this area.

In addition to the studies above, Misic & Johnson (1999) provides a better segregation in benchmarking technique according to ``functional benchmarking'' versus "competitive benchmarking". Functional benchmarking, would any identified organization that connected with the target web sites or industry studies. Unlike competitive benchmarking, where sites of direct competitors only are used, herewith, functional benchmarking considers organizations that have resemble characteristics. This functional approach does not preclude using just competitors, but a broader range of organizations.

In the next section, I shall discuss how the theoretical concepts influence my research methodology.