A User-Centred Framework for Website Evaluation

Sangeeta Karmokar
Auckland University of Technology, sangeeta.karmokar@aut.ac.nz

Harminder Singh
Auckland University of Technology, hsingh@aut.ac.nz

Felix B. Tan
Auckland University of Technology, felix.tan@aut.ac.nz

Follow this and additional works at: http://aisel.aisnet.org/confirm2013

Recommended Citation
http://aisel.aisnet.org/confirm2013/5

This material is brought to you by the International Conference on Information Resources Management (CONF-IRM) at AIS Electronic Library (AISeL). It has been accepted for inclusion in CONF-IRM 2013 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
A User-Centred Framework for Website Evaluation

Sangeeta Karmokar  
Sangeeta.karmokar@aut.ac.nz  
Auckland University of Technology

Harminder Singh  
Harminder.singh@aut.ac.nz  
Auckland University of Technology

Felix B. Tan  
Felix.tan@aut.ac.nz  
Auckland University of Technology

Abstract
The growth of the Internet has encouraged the creation of visually rich and perceptual interfaces on personal computers and mobile devices. Organisations develop websites for various purposes, and over time, the features and functions of websites have evolved significantly. Since website quality affects organisational performance, it is important to be able to assess the efficacy of websites. However, there are two key issues with the literature on website evaluation: a) a focus on specific aspects of website performance, not their overall impact, and b) limited attention on their ability to meet the broader needs of users, beyond usability and functionality, such as their social and emotional concerns. This paper uses design science to develop a theoretically-grounded evaluation framework for this purpose. Drawing on Shneiderman (1998) and Brown (1999), the framework proposes that website evaluation should triangulate information from two sources (users and experts) and using different methods (task analysis with users, in-depth interview with users and expert reviews). The framework is applied in a website development project, and the results are discussed.

Keywords
User centred design, Website Evaluation techniques, User task analysis, Expert review analysis, Customer experience

1. Introduction

Successful websites are crucial for businesses as they provide a platform for promoting their products and services and are an avenue for generating revenue by attracting more customers. Unfortunately, not all websites successfully turn visitors into customers, or retain customers after their initial purchases.
Websites are considered successful if they retain customers’ attention, provide them with the information they require, and enable them to carry out the necessary transactions (Zeithaml, Parasuraman, & Malhotra, 2002). Studies have found that features such as visual communication (Bostock & Heer, 2009), task analysis (Kules & Shneiderman, 2007), emotional usability (Kim & Moon, 1997) and user value (Boztepe, 2007) can help firms improve the satisfaction of their customers with their websites. Satisfaction, along with trust, is crucial for customer loyalty (Dick & Basu, 1994). Therefore, evaluating the effectiveness of websites has become a point of concern for practitioners and researchers (Chiou, Lin, & Perng, 2010).

Various website evaluation approaches have been introduced, focused on outcomes such as website usability and design (Agarwal & Venkatesh, 2002; Benbunan-Fich, 2001; Nielsen, 1993), content (Baloglua & Pekcanb, 2006; Robbins & Stylianou, 2003; Vitae & Huang, 2002), quality (Cao, Zhang, & Seydel, 2005; Kim & Stoel, 2004; Zhang & Dran, 2001), user acceptance (Chung & Tan, 2004; Koufaris, 2002), and user satisfaction (Devaraj, Fan, & Kohli, 2002; McKinney, Yoon, & “Marian” Zahedi, 2002). These capture aspects of a customer’s experience while interacting with a website, but are inadequate for understanding the customer’s overall experience, which occurs in specific individual, social and cultural contexts (Maguire, 2001; Mcquire, 1974). These contexts shape their expectations, judgment and loyalty, and influence their perceptions of value and service quality, consequently affecting their loyalty (McKinney et al., 2002; Nanni, 2004). There is thus, a need for a website evaluation protocol that captures the broader issues relevant to customers. In this paper, we call this a “user-centred approach” because by using triangulated method we seek to understand end-users’ needs, aspirations, and goals, and the environmental conditions and constraints in which they live (Fuller, 2007).

This paper draws on design science research (DSR) to present a framework for developing such an evaluation tool. The paper begins by discussing current website evaluation methods, and then reviews research on artefact evaluation in design science. Based on this, the paper presents a website evaluation tool that relies on triangulating feedback from users and expert designers. This tool’s viability is examined by a case study of a website development project. The paper concludes with a discussion of the study’s limitations and suggestions for future research.

2. Background

As the number of Internet users has increased, so has the number and variety of websites. Beginning as personal information dissemination channels, they are now used for a range of purposes, such as trading products and services, playing games, promoting causes, and entertainment. As they have changed, the goals have also broadened to include customer loyalty, user satisfaction, brand awareness, and trust.

However, while the objectives and features of websites have changed considerably, the ways in which they are evaluated have lagged behind. According to Ivory (2003), website design and evaluation practices have not changed dramatically between 1997 and 2002, particularly with respect to design guidelines, accessibility optimization, usability assessment, and automated design/evaluation support. Effective evaluation is becoming more important as the variety of websites increases. Lack of effective evaluation affects the quality of the website launched. Poor-quality websites with design and usability issues can lead to frustrated end-users (Fisher, Craig,
Existing website evaluation methods emphasise the interface usability perspective and assess security, reliability, functionality, flexibility and other usability-related attributes (Nielsen, Overgaard, Pedersen, & Stage, 2005; Nielsen, 1993; Norman, 2002). For example, Zhang and Dran (2000) propose a hygiene and motivator model for website design and evaluation that focuses on key considerations pertinent to interface design. Ivory (2003) investigates automated website evaluation and concludes that the development of automated website evaluation tools and methodologies is still in an infancy stage. Dhyani, Ng, & Bhowmick (2002) examined the fundamental graph characteristics relevant to website design and classify a set of important metrics for quantifying Web graph properties, such as page significance, page similarity, and usage characterization. Zhou, Chen, Shi, Zhang & Wu (2001) analyse individual visiting behaviours and browsing patterns to evaluate website link structures. Hsiao and Chou (2006), anchored in Gestalt psychology principles, measure the Gestalt-like perceptual degrees in home page design to identify the essential visual patterns that allow designers to build effective websites.

Individually, these frameworks address different aspect of consumer experiences such as usability, information design, graphical, psychological, user behaviour and motivational needs. While these are important attributes of websites, they do not address the emotional, social, behavioural and psychological needs of customers, which are part of their overall experiences. This situation reflects Lamb & Kling’s (2003) argument that the user concept is too narrowly defined in the research and practice of information systems design, development, and evaluation. By relying on individualistic models that emphasise task models, ergonomic factors, and cognitive psychodynamics, research in this domain has adopted a limited view of users.

Current website design evaluation frameworks are on focused on evaluating the functional requirements of the users and on the usability of websites, such as their appearance, navigation, functionality and interaction. The evaluation process primarily consists of asking those who ordered the artefact what exactly they want the artefact to do. There is thus a large gap between those who design and evaluated the technology and those who actually use it (Berg, 1998).

This situation points to the need for a new approach for evaluating websites that incorporates the broader needs of customers, such as their psychological, social and emotional needs, along with a focus on usability, and includes expert opinions, in addition to user feedback (Holzinger, 2005). The next section discusses how evaluation is carried out in design science research. This field is a suitable source of ideas for this study, as it offers a variety of guidelines for developing and assessing artefacts (Baskerville & Harper, 1998; Gregor & Jones, 2007; Hevner, March, Park, & Ram, 2004; March & Smith, 1995).
3. Evaluation in Design Science Research

The approach for developing artefacts in design science research (DSR) (e.g., (Baskerville & Harper, 1998; Gregor & Jones, 2007; March & Smith, 1995)) first clarifies the goals of the artefact (constructs, methods, models, or instantiations), and then builds and evaluates its utility, reliability and validity (Hevner et al., 2004). A variety of different evaluation approaches have been developed, originating from IS, management, computer science, and other allied disciplines (Cleven et al., 2009). A design science approach would place additional emphasis on the iterative construction and evaluation of research instruments, with an aim of ensuring that the instrument design is well grounded in both theory and empirical evidence to establish its validity, reliability, and practical utility (McLaren & Buijs, 2011). An artefact is relevant when it addresses a real business need, and has been rigorously developed when the pertinent theoretical foundations and methodologies have been appropriately applied (Cleven et al., 2009).

Among their seven guidelines, Hevner et al (2004) require researchers to rigorously evaluate design artefacts, and offer five kinds of evaluation methods: observational, analytical, experimental, testing, and descriptive. However, they do not provide much guidance in choosing among extant evaluation methods. March and Smith (1995) emphasize evaluation as one of the two activities in design science: build and evaluate. Beyond simply establishing that an artefact works or does not work, evaluation should also determine how and why it works (or does not) (Pries-Heje, Baskerville, & Venable, 2008).

Venable (2006) classified DSR evaluation approaches into artificial and naturalistic evaluation. Artificial evaluation evaluates a solution technology in a contrived and non-realistic way, while naturalistic evaluation explores the performance of a solution technology its real environment i.e., within the organisation. For Markus et al. (2002), design principles emerge from a search for kernel theories that would satisfy solution requirements. Kuechler and Vaishnavi (2008) consider that most design science research deals with human-artefact interaction and, as a result, evaluation takes the form of experiments. Frank (2007) suggests that the reference model used to construct the artefact be evaluated from the economic, deployment, engineering, and epistemological perspectives so as to have a holistic assessment of its quality.

**Design Science Framework for Website Evaluation**

The aim of this section is to identify a framework for evaluating a user centred website. Pries-Heje et al., (2008) proposed a framework for DSR evaluation. The framework offers a strategic view of DSR evaluation that is useful in analysing published studies, and in surfacing the evaluation opportunities that present themselves to IS DSR researchers. A strategic framework serves (at least) two purposes: it could be used to help design science researchers build strategies to evaluate their research outcomes and to achieve improved rigour in DSR.

There are several aspects that will be valuable in formulating a strategic framework for DSR evaluation. Both ex ante and ex post evaluation is useful in identifying the validity of the artefact. Ex ante evaluations take place before the system is constructed and ex post evaluations take place after the system is constructed. Evaluation can be conducted in artificial and naturalistic settings. Artificial evaluation has advantages such as more control and lower cost and on the other hand naturalistic evaluation is more realism (Pries-Heje et al., 2008).
There are two main dimensions in this framework: time and evaluation method (Figure 1). The framework answers some very important questions.

- **What** is being evaluated?
- **How** is it being evaluated?
- **When** is it being evaluated?
- **Who** is evaluating it?

“What” involves choosing between the design process and the design product (Walls *et al.*, 1992), while “how” involves selecting from naturalistic or artificial forms of evaluation (Venable, 2006). There are advantages to both artificial evaluation (such as more control and lower cost) and naturalistic evaluation (more realism). Evaluation in artificial settings is not limited to experiments, but includes imaginary or simulated settings where the technology (or its representation) can be studied under substantially artificial conditions (Pries-Heje *et al.*, 2008). “When” is an issue of conducting the evaluation from an *ex ante*, *ex post*, or both perspectives. *Ex ante* evaluations take place before the system is constructed and *ex post* evaluations take place after the system is constructed. “Who” incorporates aspects of the evaluation context, such as users and organisations.

### 4. Proposed Framework for Website Evaluation

This study was carried out as part of a project to build a new website for a New Zealand-based small business in the sports management industry. The process for designing the website included two rounds of interviews- the first one was to gather requirements from the firm’s users and management, and the second was to collect feedback from users on the website prototypes that the designer had come up with. This second round of feedback was used by the designer to develop the final design of the website. Following this, an evaluation process was carried out.

Applying Pries-Heje *et al.*’s (2008) framework to this study, the questions that it asks can be answered in the following ways:

- **What** is being evaluated? - The design product (website) was being evaluated.
• **How** is it being evaluated? - A naturalistic approach was adopted, in that the website was evaluated by potential and actual users. Expert designers were also interviewed to ask their views about the design process.

• **When** was it evaluated? - The website was evaluated ex post (after it had been developed).

• **Who** is evaluating it? - Users and designers evaluated the website.

For the “how” question, the artefact was evaluated by using three methods (Table 1). This follows Isbister’s (2006) suggestion to use multiple methods to evaluate artefacts. He recommended the use of emo cards, open-ended testing and a sensual evaluation instrument using patterns to enhance the rigour of the evaluation process, rather than just focusing on an artefact’s usability.

<table>
<thead>
<tr>
<th>Evaluation Method</th>
<th>Goal of Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Task Analysis</td>
<td>To test the usability and functionality.</td>
</tr>
<tr>
<td>In-Depth Interview</td>
<td>To determine whether users were satisfied with the integration of their needs such as psychological, social, cultural and cognitive based on Brown’s (1999) and Shneiderman’s (1998).</td>
</tr>
<tr>
<td>Expert Review Analysis</td>
<td>To review the proposed evaluation framework (social, cultural, psychological and cognitive needs) in their practice.</td>
</tr>
<tr>
<td>1. Survey</td>
<td></td>
</tr>
<tr>
<td>2. Interview</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.** Evaluation Methods Used in this Study

First, user task analysis was used to evaluate the usability of the website (Ahmed, McKnight, & Oppenheim, 2006; Al-Qaimari, 2007; Hackos & Redish, 1998; Nielsen, 1993; Wright & Monk, 1998). Second, in-depth interviews with users based on multidisciplinary principles (Brown, 1999; Shneiderman, 1998) were used to assess whether their social, cognitive, psychological, emotional, visual/graphical and emotional needs were addressed in the website. In-depth interviews are frequently used when working with multidisciplinary principles (Al-Qaimari, 2007; Chang, 2006; Nesbitt, 2005a, 2005b). Third, a group of experts were surveyed and interviewed to gather their views on the design process. These methods are described in more detail below.

**4.1. User Task Analysis- Think Aloud**

Participants were randomly selected from a sample of students and parents for this evaluation phase. Participants were encouraged to think out loud as they work on the task. Prompting and echoing was used to encourage participants to think out loud. The tasks were selected based on the expectations of what the website’s users would accomplish through it. Each participant was asked to perform five tasks. The tasks were selected based on the expectations of what the website users would accomplish through the sites. The results were categorised in terms of their match with users’ psychological, cognitive, trust and emotional needs.
Psychological Analysis

Psychological analysis shows how users react or how they are going to behave when using online information. Users have limited time so that if they invest time in entertainment, they tend to demand more intense, more concentrated, and more satisfying returns. Users usually relate what they encounter to what they have seen and assume that their past experiences are the norm for all online information. For instance, they try to click on an image assuming that is a video or a hyperlink. Users respond fast when graphics are used compared to the text. It is much easier to communicate with users visually.

Cognitive Analysis

Cognitive skills include language, memory, perception, learning and development and attention. Cognitive analysis shows that not many users read the entire information on the website. They scanned or glanced at the information provided on the website and tried to relate with the words. Users do not like being distracted by animation when they are trying to process mentally and correlate with their hand movements. It is difficult for users to perform multiple things on the screen. Uses and gratification theory approach posits that when e-business employs techniques that are too flashy with big-size graphics or abuses those techniques by tracking consumer information and behaviour online, Web users may perceive this as an unwanted, offending, and negative influence (Suh, et al., 2010).

Emotional and Trust Analysis

Emotional analysis shows that the websites need to be strategically designed to address the users concerns. Home sections is very important for users, it is the place from where they connect with other sub sections of the website. Going back to the homepage makes the user feel safe and secure and provides reassurance that they are in the right place. Users felt emotional satisfaction when they saw personal videos and blogs. Users create the content of these sections and provide a feeling of expressing themselves to others. The information and achievements of the organisation created a trust that it is reputed organisation. Websites interactive elements such as e-mail, instant message, chat room, message boards, and other Internet venues created a feeling of trust towards the website.

4.2. In-depth Interviews

The users who participated in the initial stage of the design process were interviewed. They had been asked about their psychological, social, personal and cognitive needs vis-à-vis website use during the website design phase. Being part of the design process made them feel that their concerns relating to trust, safety, confidence, reassurance and personal values had been integrated into the final layout.

Users felt safe and confident when one or more options were provided to navigate, because that created a feeling of trust and security and they did not have to fear getting lost on the website. Providing information about the organisation and the individuals involved in the organisation created a feeling of personal bonding with the website. This made users feel assured and trustworthy about the organisation and the content. Personal information on the website made it more personalised and added a layer of credibility. The stylish look of the website enhanced the organisation’s social status, and made them feel that it had high business standards. Providing communication and social networking tools gave them a sense of belongingness. Users perceived
a connection with the target market and felt that the organisation was establishing social bonds 
with the users.

The website was able to create trust and positive perception about the organisation. Users felt a 
sense of belonging to the website as they saw that their views had been integrated in the final 
layout.

4.3. Expert Review Analysis

The expert review was conducted using two methods: survey and interview. Surveys are cost 
effective, less time consuming and have higher response quality, while interviews provide more 
detailed information than what is available through surveys. Combining both methods provides 
an in-depth understanding of the expert’s views.

Ten expert designers were selected from the software design, website design and design 
academic fields. While a single evaluator could miss out many problems, different evaluators 
will find varied problems. Thus, better results can be obtained by combining the results from 
several evaluators.

An evaluation matrix was used to collect the feedback from the designers. This was based on the 
criteria provided by researchers for evaluating user-centred design (Abras, Maloney-Krichmar, & 
Preece, 2004; Gulliksen et al., 2003; C. M. Johnson, Johnson, & Zhang, 2005; Plass, 1998) from 
cognitive (Johnson et al., 2005; Spillers, 2001), psychological (Kramer, Noronha, & Verga, 
2000), emotional (Roy, Dewit, & Aubert, 2001; Spillers, 2001), usability (Nielsen, 1993; 
Norman, 2002; Preece, 1994), and information design (Jacko, 2007; Shedroff, 1999) disciplines.

The designers strongly agreed that the new evaluation process integrated all the needs of users in 
final design, including usability. They felt that the new evaluation process was intensive and that 
there were not many chances for the designer to redo the design or fix any problems at a later 
stage after implementation.

Overall, the results indicated that the new evaluation process integrates usability and the broader 
needs of the users. The process allows designers to look at both the users’ and clients’ needs in a 
less complicated way.

5. Conclusions

The paper provides an alternative approach to website evaluation, based on the evaluation 
literature from design science research. The new triangulated evaluation method uses data from 
multiple sources (users and experts), some of whom were involved in the artefact’s design while 
others were not, and multiple instruments (interviews, surveys, and task analyses). This enabled 
the evaluation framework to measure user performance and satisfaction, and collect expert 
feedback.

This evaluation framework is an example of a user-centred approach, as it supports users’ needs 
and their information-seeking behaviour. Until designers apply such evaluation techniques, most 
interface designs will be driven by the constructional domain such as engineering, and possibly 
by computer scientists, rather than by the needs of the users for whom these interfaces are 
intended. We hope this work provides a starting point for techniques that let researchers and 
practitioners evaluate interfaces that are both easy to learn, use and remember.
We are excited about the possibilities for the evaluation framework to be used in further enabling design teams to engage in productive and reasonably scaled user testing that improves the emotional experience for end users. The evaluation framework is also applicable to processes that are not necessarily technology-based. For example, it can be applied to processes where users are vital parts of the artefact, such as business processes, service innovations, and maturity models.

References


Spillers, F. (2001). Emotion as a Cognitive Artifact and the Design Implications for Products that are Perceived As Pleasurable. Experience Dynamics.


