User Experience: Assessing the Effectiveness of Internet Booking Service

Idyawati Hussein
Universiti Malaysia Sabah
idyawati@ums.edu.my

Murni Mahmud
International Islamic University Malaysia
murni@iium.edu.my

Abu Osman Md Tap
International Islamic University Malaysia
abuosman@iium.edu.my

Abstract

Despite the undeniable importance of usability and utility, one should not forget that people use Internet booking service for variety of reasons or in essence overlooked the vital intent. Online booking systems challenge the effectiveness of traditional way of booking process effectively. Therefore the effectiveness of the task is crucial to support the user to complete the tasks. User Experience (UX) is basically a constant “good-bad” feeling towards any products or service to finish a task. Standard heuristics evaluation has been performed in UX to measure if the perceived usability of the online booking based on the perceived utility matched to the user’s satisfaction. This paper indicates that several design features do not entirely conform to the usability standard and design principles. The finding shows that the users experience some difficulties in completing the task; however, they proceed in order to get competitive prices. They also attribute the difficulties to their incautious and negligent without realizing the design features have lead them to experience human error in the online booking system.

Keyword: User Experience (UX), heuristics evaluation, Internet booking service, user interface, Malaysia

1. Introduction

A product’s ability to satisfy what people have to do with ease is a matter of product usability and utility – quality aspects at the heart of practical and academic Human–Computer Interaction (HCI) [7]. ISO CD 9241-210 tentatively defines User Experience (UX) as “all aspects of the user’s experience when interacting with the product, service, environment or facility. It includes all aspects of usability and desirability of a product, system or service from the user’s perspective” [10]. On the other hands, researcher described UX definition in two parts: one that defines UX itself and a second that states how UX is “made” [7]. Evaluation of websites has adopted two approaches: first, observation of users’ errors when navigating websites; and secondly, expert style heuristic evaluation in which the quality of the interface is judged against a set of criteria-the heuristics [16].

In recent years, there has been great excitement about using the Internet as the most effective platform to automate task and reduce merchandising costs [12]. Asian companies, including Malaysian companies, are responding to this new reality. While companies have embarked on transferring merchandising skills to the web site, issues of the website user interface design effectiveness have rarely been addressed or evaluated [20].

User interface for business websites affects many features of peoples’ lives as it is part of the system/product with which people come into contact physically, perceptually and conceptually [2]. The lack of vision for the user is they have lost the opportunity to voice what is wrong with the current user interface by mistake; not knowing what is good or bad design. To assure that users are in essence design educated and guarantee that user interface design conforms to ergonomic design universal, evaluation of the effectiveness of user interface is needed [5]. Whether the
evaluation is done in a formal setting or a self-evaluation (self-talk); both could perhaps be beneficial for the end users. This paper presents a heuristics evaluation in a formal lab setting to identify users’ satisfactions in using online booking system.

This paper is motivated by two problems: first, how definitions of effectiveness need to be examined in light of experience of the web applications; and how human-error can be prevented in terms of design guidance. Secondly, how reliable heuristics judgment towards observing users’ errors when completing a given task on website is. The paper is structured in three sections. Firstly, literature on UX and heuristics evaluation is presented for websites and e-commerce. This is followed by a case study that assesses the reliability of heuristics for evaluating the usability and safety of Internet booking service. This paper concludes with a brief result and discussion.

2. Literature Review

In practice, product development teams may extend systematic evaluation to real usage because at this stage the web interface is on the market and is not very difficult to fix [20]. Thus, heuristics evaluation methods have been extended for web interfaces [16]. HCI-oriented sought the most reliable UX evaluation data comes from people who have actually purchased and used a product on the market [19].

In the context of HCI and UX, [8] suggested that interactive products are perceived by their users/owners with regard to their capability to fulfill what users want to do (i.e. online booking) and what users’ need (i.e. promotion price according to their affordance)[7]. Studies [6] show that individual’s motivational orientation influence product evaluation and choice. Thus, people may perceive products promotional as primarily important than perceived value of interactive products. In addition, products appeal (e.g. websites that lead to mistakes) and choice is strongly context-dependent.

To date, research have shown that results of objective and subjective evaluation correlated poorly due to the character of users in the study; subjects were less vocal, exceedingly polite and disinclined to express negative comments in front of observers [12]. Although Malaysia has implemented web interfaces in various field, the effectiveness, efficiency and satisfactory of these web interfaces have rarely been evaluated or published [11] and few comparisons have been made [1]. Research has shown that one of the reason for this lack of study in the area of user interface in Asia could be caused by the Asian culture whereby “it is considered culturally unacceptable to criticize the designer directly, as this may cause the designers to lose face” [9]. Malaysians try to refrain from giving negative comments or feedback of user interface design in order not to degrade the designer [1]. A study in human-computer interaction related field in the region also found that user broke down and cried during a software evaluation session [9]. Consequently, it seems crucial to monitor product experience throughout the whole product lifecycle and to use these findings at least as a guide for future products design.

Recent work by Malaysian researchers draws on the new paradigm of producing desirable [12] websites as opposed to concentration on website usability and performance [4]. Even if the basic needs of the safety (e.g. error avoidance principle on user interfaces) are fail to be observed in the design, concentration on engagement and fun will inevitable fail to be produced. Studies of website design evaluation based on cross-cultural issues is important step towards adopting Grounded Theories in social science to further investigate Malaysian website user interface design [4][16]. [12] develop an evaluation tool for usability; however, the usefulness of such tool has never been refined and tested by
industry. Appropriate publicity and awareness should be provided to persuade people to support HCI at every stage of design, development and deployment through public demonstrations, media, word-of-mouth and other mechanism.

3. Method

The lab evaluation involved the experiences of different users applying Nielsen heuristics in UX evaluation methods and highlighting their awareness of the online booking task. Measures that can be obtained from the evaluation include number of violations of guidelines or heuristics [3].

3.1. Demographics

Each data collection session consisted of thirty (N=30) users and lasted approximately 4 hours. The subject population was composed of undergraduate, postgraduate and staff of the Universiti Malaysia Sabah. The users were chosen as representative of the online booking website users, particularly as the ‘expert’ because all of them have experienced using neither nor both websites for online booking. The sample was 33% male and 67% female as shown in Table 1.

Table 1. Demographics Information of Evaluators

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>10 (33%)</td>
</tr>
<tr>
<td>Female</td>
<td>20 (67%)</td>
</tr>
</tbody>
</table>

About 94% of user reported that they spent more than 2 hours on the Internet per day, 3% spent less than two hours per day and 3% spent at least one time browsing the Internet within three days. Users have indicated that they use computer for an average of 9.03 years. Figure 1 shows the demographics information of the users. 71% of users were Malay, 11% Kadazan, 3% Kadayan, 3% Orang Asli Suku Semelai, 3% Irranun, 3% Bajau, 3% Bidayuh and 3% Melanau.

3.2. Procedures

The evaluation took place in a laboratory which was connected to a 100.0 MB per second internet speed. At the beginning of the experimental session, users received written and verbal instructions and followed by pre-test questionnaire to record personal data. Users were then, asked to browse the website to familiarize themselves with the websites’ content. While browsing, users were asked to report their impressions of the website, including thoughts on content, presentation and usability issues in section one of the questionnaire.

Next, users were informed that they would be working in pairs for section two of the questionnaire. Each pairs was given a task. One user was directed to perform the task and another user was recording the time performed on each task. Afterwards, users repeated the same procedure for the second website. After exploring both websites, users were then asked to discuss on each heuristics and reported their feedback.

4. Result and Discussion

The usability was assessed by objective measures (performance and self-report of usability problem) and subjective measures (questionnaire). Each evaluator was asked to work in pairs for this session. 15 pairs evaluated the performance of both websites. The average of completion time for A is 7 min 29s and 4 min 57s for B. Each pair was
asked to complete one task and follow the booking scenario.

The users rated each site on a 1-5 scale for each heuristic and were asked to report rationale of their decisions and the ease with each heuristic could be interpreted. The rating scores were converted into Net Positive Value (NPV) [17] to reflect the range of the evaluator’s assessments. The frequency of the users’ ratings is multiplied by the +2 to -2 scale and the product is totaled to give a value for the heuristic. Table 2 shows a worked example of the NPV.

Table 2. A Worked Example of Net Positive Value (NPV)

<table>
<thead>
<tr>
<th>Attractiveness Rating (1-5)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Rating Frequency (subjects)</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Product (Scale x Rating Frequency)</td>
<td>-2</td>
<td>-3</td>
<td>0</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Total Net Positive Value (NPV) = (Sum of Product)</td>
<td>(-2)+(-3)+0+11+10= +16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In figure 1, both Internet booking service websites aim to provide promotional price information as well as other online service available. The ratings of each site are given following the model as a cognitive walkthrough with a common scenario of buying a ticket, which is similar to previous work [17]. The overall assessment of the sites is given in Table 3.

Table 3. NPV scores for each site on attractiveness, persuasion and usability.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall attractiveness</td>
<td>16</td>
<td>43*</td>
</tr>
<tr>
<td>Persuade to buy</td>
<td>40*</td>
<td>18</td>
</tr>
<tr>
<td>Usability and Navigation</td>
<td>24</td>
<td>25*</td>
</tr>
</tbody>
</table>

Overall, B received higher scores in attractiveness (NPV=45) as compared to A (NPV=16). B website has a background in half shades and low saturation color which provides a more depth and interest in an image [20]. The uses of dynamic media of realistic images and animations have improved the attractiveness of the e-commerce website presentation. However, the media selection is suffered from considerable bandwidth and Internet speed. A scored lower in attractiveness due to the use of intense red colour whereby limits the usage of foreground image selection. For overall attractiveness, 11 (36.66%) users commented that both online booking websites have fulfilled the aesthetically pleasing design; unfortunately, the issue of usability needs to be addressed.

A website is more persuasive (NPV=40) than B (NPV=18). According to the more than 50% of the users, the label of well-recognized certification (e.g. “World’s Best Low Cost Airline” and “verified secured”) has a high impact on their impression and attitude to purchase online booking (reported by 9 users). Three (3) users concluded that although A advertises its Internet booking extensively, they rather choose B because they trust the company owns the website. Both websites scored about the same in the overall usability and navigation (A NPV=24, B NPV=25).

B was rated more favorably than A on well-structured pages; however, several evaluators reported that B website is lack of terminology consistency for example in command naming such as “logon” instead of login. When asked about their feelings of the terminology; all (100%) Kadazan users, Kadayan, Orang Asli Suku Semelai, Irranun, Bajau, Bidayuh and Melinau found that the word is alien to them. However, these evaluators blame themselves for being technology inadequate. Figure 2 shows an
example of B website violation in the internal consistency principle.

Internal consistency is the degree to which the same appearance, meaning and operation of interface attributes hold within the same application [18, p203]. The use of “Logon” terminology creates awkward feelings amongst other ethnics except Malay. This may be due to the Malay ethnic originally from West Malaysia who uses Internet booking service more than local users (originally from Sabah or Sarawak region).

Table 4 illustrated the result of Nielsen’s heuristics assessment of both websites.

<table>
<thead>
<tr>
<th>Table 4: NPV for Nielsen Heuristics Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visibility on system status</td>
</tr>
<tr>
<td>Match between system and the real world</td>
</tr>
<tr>
<td>User control and freedom</td>
</tr>
<tr>
<td>Consistency and standard</td>
</tr>
<tr>
<td>Error prevention</td>
</tr>
<tr>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>Help users recognize, diagnose and recover</td>
</tr>
<tr>
<td>From errors</td>
</tr>
</tbody>
</table>

B website scored slightly higher on visibility of the system status (NPV=27) notwithstanding that A website supply obvious site maps and consistent navigation menus (NPV=20). A and B being the same on the match between system and the real world (A NPV=14, B NPV=13) and scored the about same on user control and freedom (A NPV=3, B NPV=7). Eighteen (18) users reported that B has a simpler form, no scroll down page and provides previous button on each page. A (NPV=26) scored higher than B (NPV=7) in consistency and standard. B scored lower than A in consistency. As can be seen from the result, A scored negative value (NPV= -13) for error prevention rule and help users recognize, diagnose and recover from errors (NPV= -4).

Figure 3 shows the example of how the principle of visibility violated in A website. [14] highlighted that "The division occurs at the level of the intention: A person establishes an intention to act. If the intention is not appropriate, this is a mistake. If the action is not what was intended, this is a slip." About 80% of the evaluators commented that they made mistakes the first time using the website because they did not see the importance of selecting additional service until the process of finalizing transaction. At that time, they blamed themselves for not changing the options for luggage, which cost them additional fees, which affected those who do not need it or need more luggage quantity. A do not have any options to go back to any chosen page, therefore they have to re-do the booking process all over again. Previous or back button is not available on the website; hence users must put the right information in the first place.

The experiment shows the presentation of dialogue, which requires a little effort in working memory. The font size and colour is too small (commented by 7 evaluators). The users think, the information should be immediately apparent. In the study of design for everyday things, the important things to watch should be visible and clearly marked; the results of any action should be immediately apparent [14].

Figure 4 shows the "Continue without picking a seat" means to cancel the action of picking any seat.
In figure 4, the system’s flow chart constructs a mental model of not picking any seat if pushing “Continue without picking a seat” button. Unfortunately, when user clicks on “Continue without picking a seat”, the raised “OK” dialogue box message still invites user to push the button and make mistake. Twenty-one (21) evaluators commented that they thought “OK” means continue without picking a seat. In this example, the website disobeys design principle “Do not use double negative as they can be ambiguous” [5, p389].

Figure 5. The raised “OK” alluded to double negatives of dialogue

The result in figure 5 found an inability of the website to formulate an appropriate mental model for users to proceed with task picking a seat [5, p389]. According to error avoidance design guidelines, if something must change, make it large and obvious [16]. A website failed to conform to this rule. Nineteen (19) evaluators found that A website violated design for error avoidance guidelines. For example an additional price for insurance option for one-way travel, which most travelers did not need. However, the system selected this option by default.

Users have to recall the steps that they must tick or choose to remove the selected option. At first, all of the evaluators thought they were careless and always blamed themselves for making the same mistakes many times. During interviewed, 83% of the users claimed that they did not see and/or missed the “cancel” insurance task although they firstly intended to remove it. The visibility to cancel additional service, which leads user to pay higher price, is violated. Likewise, users have to make effort to move to the word “cancel” to know if it is clickable as can be seen in figure 6.

Another example of short-term memory violation in this study was when user clicks on cancel word. Yes-No dialogue boxes will pop-up. The user needs to resort to previous (nonafforded) knowledge about what option should reply to “cancel”. With limited time and working memory, user was invited to click “OK” – thinking that it means “OK” to cancel insurance; however, it means the opposite.

Figure 7 shows the pop-up window of dialogue message after user clicks on cancel. If user do not cancel the insurance and proceed to the next task, user cannot easily change it. As a result, users have to pay additional fee for the service and to change it, user need to do the whole booking process again.

During the lab think aloud session, all nine (9) evaluators who had never attended a formal HCI class commented that A website is lack of honesty. The principle of honesty relates to the ability of the user interface to provide an observable and informative account of a change for example insurance policy. In A website, much critical information was carefully hidden steering to users’ mistake or slip [2] [14] which impacted users to pay more than the first intended price.

5. Conclusion

The contributions of this paper have been to expose the ineffectiveness of design element features using heuristics evaluation,
which induce to users’ slips. According to [14], "slips are most likely to occur (a) when we must deviate from a routine, and automatic processes inappropriately override intentional, controlled processes; or (b) when automatic processes are interrupted - usually as a result of external events or data, but sometimes as a result of internal events, such as highly distracting thoughts". The results demonstrate that many public users were unaware or pay no heed to manipulative designs. Despite, they blamed themselves for being technology inadequate and still can tolerate the price of the product or service, which was considered being “still cheap” as compared to manual booking. After participating with evaluation study, 100% users agree that HCI knowledge such as poor design interfaces, evaluation methods, ergonomic standards and principles are very important.

The knowledge must be exposed to Malaysian public and community through education and any social media available. Public users must be warned to put on a constant stream of self-talk while they experience the interactive products, which are in related to human-product interaction. The constant “good-bad” feelings should always be part of the experience but must not necessarily do so [5]. Business stakeholders, software developers and designers should incorporate fitness for purpose design, which support user’s mental capabilities and apply the ethical use of design features. Considering the findings, it is important for practitioners and researchers in the field of HCI and user-centered design to create awareness among Malaysian users that interface design must be evaluated and tested by real users.

The contributions also have been to describe heuristics evaluation method, which is based on an experimental investigation in the light of UX currently being conducted in a Malaysian university lab. Although, [16] concluded that heuristics should not be used for subjective rating style judgment, this study has diagnosed ad towards Malaysian Internet booking service usability evaluation which was Nielsen’s original intention [13]. Future studies could take on UX based on values, emotions and context-dependent [7][9][12][16].

6. References


