Designing and conducting an empirical experiment comparing two web sites

Wafaa S. AL-Baldawi

Computer Science Department, College of Science, University of Kerbala, Karbala 56001, Iraq
wafaashamh@gmail.com

Abstract. This paper seeks to present a usability evaluation of the two websites (www.bmw.co.uk, www.ford.co.uk) by performing three tasks related with ability to access, navigate, and find the information to complete the tasks without using the search engines with answering some paper-pencil questions after each task. A sample of 16 users is used to collect the actual data. There are three types of questions have been asked including open ended, closed-ended and partially closed items. While the user is doing the task, many factors have been measured to determine which website is more usable than other. Designing the empirical experiment is to identify the major strengths and weaknesses of the site and to incorporate the results and user feedback.

Keywords: Website usability, satisfaction, effectiveness, efficiency.

1. Introduction
This paper reported to compare two websites (www.bmw.co.uk, www.ford.co.uk) in term of usability by performing the three tasks related with ability to access, navigate, and find the information to complete the tasks without using the search engines with answering some paper-pencil questions after completed each task by participating the actual users.

The usability can be defined as the effectiveness, efficiency and user satisfaction by determining if a sample set of users can complete the specific set of tasks in a lab environment [1].

When the user is doing the task, the number of errors, time taken to complete a specific task, verbal protocol, visual protocol, visual scanning patterns, and attitude had been measured to know which website is easier to use than another. Also, if the users satisfy with a specific website separately and if the task can be completed or uncompleted in a formal laboratory experiment. The goals of the evaluation had been defined by identifying the objectives of an evaluation to make sure the methods which should be used and the factors that should be measured are considered [2].

Three evaluation measurements are used: the efficiency, the effectiveness, and the users’ satisfaction. These aspects illustrates in the following:

1. The efficiency:
In evaluating the usability of a website, it is important to calculate the time required to perform a specified task as well as the number of errors encountered to retrieve a particular information and their typologies [3][4][5]. Badre [5] mentioned that errors are the hardest decision in the usability measures and that includes identifying the sort of errors happening and there frequency of occurrence. To solve this issue, the experimenter can compute the misdirected or incorrect activities that will not let the user to complete the requested task.
2. The effectiveness of the website:

Navigation can negatively affect the usability of any website and this can be related to the web-user’s expectations, and the positioning and the availability of the controls in the specified page [5].

Krug states that many experts in the usability and designing websites waste a lot of time calculating and relating the numbers of clicks required to get the desired information without irritating the user. Even set a maximum number of clicks that should not be exceeded to get a piece of information when designing the website [6]. The concentrate should be on how easily the navigation design (such as content structure and adequacy) is used to allow the user to obtain the information demanded [7].

3. The Users’ Satisfaction:

Collecting, and analyzing self-reported metrics such as user satisfaction, attitude, and feelings about the explored website is necessary to complete the overall evaluation of any website since high performance measurements, in many cases, do not mean positive satisfaction (vice versa) [8]. Good user satisfaction results are what a website requires. These will gain the users’ attention and frequent visiting [9]. These subjective/qualitative metrics may be found by asking users carrying out survey techniques after finishing the experiment [7].

The organization of this paper is as the following: section 2 introduces research model and hypotheses that includes users, experimental design, variables, apparatus and materials, description of user interaction, procedure, discussion and conclusions, and limitation and recommendations. In section 3 references are listed. Finally, appendices are the section 4 which includes: tasks’ sheet, questionnaire that includes two parts: questionnaire for users' background and users' subjective opinions questions.

2. Research Model and Hypotheses

To test the usability, designers should collect the qualitative and quantitative data from real users. The final basis of this work is depending on what in the literature review. Effectiveness, efficiency, and satisfaction are proposed the three first-order factors usability construct. Thus, the following hypotheses are posited:

H0a: There is no difference between Ford website and BMW website in navigation.
H0b: There is no significant difference in the effectiveness level of user interaction in the BMW website and the Ford websites.
H0c: The ford website is the same as the BMW website in terms of user satisfaction.

2.1. Users / Respondents

Users must meet the following criteria in order to participate:

Users are selected from different nationalities who are experienced driver with a valid driving license as the website to be tested is aiming to attract driver.

1. Not sober, not drunk and not drug trials.
2. The age more than 18.
3. Drivers and/or interested with cars.
4. Do not work in car companies or as website designers because they may be working as the designers of the websites for competitors companies.

The number of users is difficult to determine and the sample of population which are available but the truly representative sample should use rather than from the population at large [10]. In this study, the 16 users are observed for testing the websites. They are undergraduate, postgraduate students and graduated workers. Users are from different nationalities and their ages between 20 and 40. They are 5 females and 11 males.

2.2. Experimental Design
The majority of website evaluation tests are carried out using comparative research. In such research a new website is compared with its old version or a website is compared with another. In this later case, the two websites are compared to find which one is more usable than another. The two websites are www.BMW.co.uk and www.ford.co.uk. There are Within subjects design and Between subjects design can be used.

The Within subjects design is preferred because the number of users is fewer than the number of users in Between subjects design. Despite the more testing is needed in within subject for each user. There is a benefit from using the within subjects such as briefing, practicing, scheduling, and so on are easier than in between users design.

Also, the variant in measurements is more acceptable due to the differences between the websites than to behavioral differences among users while using the Between-subjects design. However, in Within subjects, the users learn something while they are testing in the first website. Therefore, using a Latin square which is a design in each test is assigned to each time period the same number of times and to each subject the same number of times [11]. Latin square is used by dividing the user in two groups and presenting the two websites in different order [12]. While the users are performing the tasks, the BMW website is presented first to Group1 then Ford website. In contrast, the Ford website is presented first to Group2 then the BMW website.

The users are treated by the experimenter in the same way and in the same environment but not all the users are tested in the same day. The time of the tests which are taken depended on the free time of each user separately. The experimenter controlled the situation of the session. However, the Internet speed and the network traffic would significantly affect the results.

The time for each usability study session is 50 minutes in which the evaluation test is carried out by each user and the quantities and qualities performances are measured.

2.3. Variables
In this study, two independent variables are considered throughout two set of tasks:
• In the first task, the user has to find a certain page and the extract details of a particular dealer.
• In the second task, the user has to check the specific feature of the exact car model in each website. The feedback is collected by the help of questionnaire. The dependent variables came in use to measure the efficiency and the effectiveness of the website. A number of errors that occurred during the entire surfing of website are calculated (such as accessing wrong pages, etc.) at the end of the whole experiment by the experimenter with the help of software (For details, please refer to ‘Apparatus and Materials’ Section).

The method is used to calculate the efficiency, effectiveness and user satisfaction. These are achieved by carrying out a questionnaire, by which users are asked to share their perception by scoring the question.

The user’s satisfaction mode can be judged on the basis of their questionnaire scores, and the experimenter’s observation. The confounding variables may become a hurdle in the middle of the experiment. The experiment is conducted in an entirely separate room to avoid all the means of distraction to the user.

2.4. Apparatus and Materials
1. The equipment uses in the experiment is a laptop running Windows 7, 3GB RAM, Intel core i7 processor with an internet connection (Wi-Fi internet connection) to surf both the websites named bmw.co.uk and ford.co.uk. The websites are surfed in the Google Chrome browser.
2. A built-in laptop camera.
3. Task based Testing summary sheet is needed for measuring the tasks achievement (see Appendix).
4. Microsoft Excel 2007
5. Digital stopwatch is used for time measurement.
6. To reduce the environment effects, the user is asked to set the mobile to the silent mode, and a water bottle and beverages/refreshments are arranged for the user in case of requirement.
2.5. Description of User Interaction
At the beginning of the test, the experimenter welcomes the user with an introduction about the objectives and the aim of the session, and provides the user with one instruction paper to guide them in completing all the activities (doing the tasks and answering the questionnaire). When the session starts, the experimenter gives the user one conform form to fill. While the user fills the form, the experimenter prepares the tasks cards, the questionnaire papers, and pens to be ready. During the session, the experimenter writes some notes, time which is taken to finish each completed task, and notes on tasks that are uncompleted and the errors encountered. The user is invited to visit both websites one after another to accomplish the two tasks. When both tasks are completed, a questionnaire is provided to the user. The main focus of preparing questions is to evaluate the efficiency of website, the effectiveness of website, and the user’s satisfaction. Above all, a guideline sheet is also introduced to the user, describing the two tasks.

2.6. Procedure
One experiment conducted by 16 users by visiting the websites and filling the questionnaire in the end of each task. The total time spent of each experiment is set.

On seeking the permission of the user verbally and through consent form as well, the experiment started. The laptop with the network connection is presented to the user. Then, the experimenter starts the stopwatch. Throughout the task, the experimenter times each task, observes the user’s activities, and notes down the result on a sheet of paper, already prepared by the experimenter.

This criterion is followed after each sub task. Once each task is accomplished, the user is given a questionnaire, containing a set of questions organized in category wise (please refer to Appendix for details of the questionnaire).

2.7. Experimental Results (Statistical Analysis)
• Statistical results of the efficiency of the two websites:

| BMW | Ford | Diff | t-Test: Paired Two Sample for Means | 0.05 |
|-----|------|------|------------------------------------|------|
| 02:02.0 | 02:03.5 |      |                                    |      |
| 06:00.0 | 03:49.6 | 02:10.4 |                                    |      |
| 03:30.7 | 01:00.0 | 01:40.7 Mean | 0.001754 | 0.001567 |
| 03:20.0 | 04:30.0 | 04:10.0 |                                    |      |
| 05:00.2 | 06:03.9 | 04:03.7 |                                    |      |
| 03:00.3 | 03:54.1 | 03:54.1 | Pearson Correlation | 0.747011 |
| 02:08.0 | 00:23.0 | 01:45.0 | Hypothesized Mean Difference | 0.00 |
| 00:54.0 | 01:00.0 | 01:46.0 | df | 15 |
| 01:59.0 | 01:00.0 | 00:59.0 | df | 0.00 |
| 00:30.0 | 01:47.0 | 04:17.0 | P(T<=t) one-tail | 0.180 |
| 01:00.0 | 01:05.0 | 01:05.0 | T Critical one-tail | 1.763 |
| 01:36.0 | 00:25.0 | 01:11.0 | P(T<=t) two-tail | 0.360 |
| 00:10.0 | 00:31.0 | 00:29.0 | T Critical Two-tail | 2.131 |
| 04:15.0 | 02:62.4 | 01:23.2 | The two values of p are greater than 0.05 so the Null Hypothesis is accepted |
| 02:05.0 | 02:50.0 | 00:45.0 |              |
| 02:00.0 | 02:01.9 |      |              |

**Figure 1.** Statistical Result of the Time Spent for Task1
### Figure 2. Statistical Result of the Time Spent for Task2

| BMW | Ford | Diff | t-Test: Paired Two Sample for Means | α | 0.05 |
|-----|------|------|-------------------------------------|---|------|
| 06:45.2 | 05:06.0 | 01:39.2 |                                      |   |      |
| 02:37.0 | 06:41.9 |           |                                       |   |      |
| 08:00.0 | 03:31.7 | 04:28.3 | Mean                                 |   |      |
| 01:59.9 | 08:00.1 | 01:59.9 | Variance                             |   |      |
| 01:58.6 | 01:55.6 | 01:58.6 | Observations                          |   |      |
| 00:05.0 | 05:01.7 | 04:56.2 | Pearson Correlation                   |   |      |
| 0.41419 |          |         | Hypothesized Mean Difference          |   |      |
| 0           |          |         | df                                   |   |      |
| 15           |          |         |                                        |   |      |
| 01:29.0 | 02:30.0 |           |                                        |   |      |
| 03:37.0 | 03:46.0 | 00:09.0 | p(T<|t|) one-tail                       | 0.421 | Accept Null Hypothesis because p > 0.05 (Means are the same) |
| 02:15.0 | 06:18.0 |           |                                        | 1.753 | Accept Null Hypothesis because p > 0.05 (Means are the same) |
| 03:22.0 | 03:09.0 | 00:13.0 | t Critical one-tail                   | 0.842 |                                        |
| 01:25.0 | 02:15.0 |           |                                        | 2.131 |                                        |
| 09:17.4 | 06:19.3 | 02:58.0 | The two values of p are greater than 0.05 so the Null Hypothesis is accepted |           | |
| 02:30.0 | 02:50.0 |           |                                        |           | |
| 05:00.6 | 04:55.0 | 00:05.6 |                                        |           | |

### Figure 3. Statistical Result of the Time Spent for Task3

| BMW | Ford | Diff | t-Test: Paired Two Sample for Means | α | 0.05 |
|-----|------|------|-------------------------------------|---|------|
| 06:51.3 | 04:10.0 | 02:41.3 |                                      |   |      |
| 05:31.0 | 05:24.0 | 00:07.0 |                                        |   |      |
| 06:45.0 | 03:10.6 | 03:34.4 | Mean                                 |   |      |
| 02:03.0 | 04:00.3 | 02:03.0 | Variance                             |   |      |
| 01:09.0 | 01:09.0 | 01:09.0 | Observations                          |   |      |
| 06:37.0 | 08:29.1 | 01:42.0 | Pearson Correlation                   |   |      |
| 0.73408 |          |         | Hypothesized Mean Difference          |   |      |
| 0           |          |         | df                                   |   |      |
| 15           |          |         |                                        |   |      |
| 03:00.0 | 02:10.0 |           |                                        |   |      |
| 02:02.0 | 03:12.0 | 00:10.0 | p(T<|t|) one-tail                       | 0.367 | Accept Null Hypothesis because p > 0.05 (Means are the same) |
| 00:01.0 | 00:53.0 |           |                                        | 1.753 | Accept Null Hypothesis because p > 0.05 (Means are the same) |
| 02:19.0 | 02:33.0 | 00:14.0 | t Critical one-tail                   | 0.734 |                                        |
| 02:00.0 | 01:45.0 |           |                                        | 2.131 |                                        |
| 04:20.0 | 04:11.0 | 00:09.0 | The two values of p are greater than 0.05 so the Null Hypothesis is accepted |           | |
| 03:00.0 | 03:00.0 |           |                                        |           | |
| 02:08.0 | 02:30.0 |           |                                        |           | |

### Figure 4. Statistical Result of the Number of Errors in Task1

| BMW | Ford | Diff | t-Test: Paired Two Sample for Means | α | 0.05 |
|-----|------|------|-------------------------------------|---|------|
| 0   | 0    | 2    |                                      |   |      |
| 2   | 0    | 0    | Mean                                 | 0.625 | 0.3125 |
| 2   | 0    | 0    | Variance                             | 0.78333 | 0.3625 |
| 1   | 0    | 1    | Observations                          | 16 | 16 |
| 1   | 0    | 0    | Pearson Correlation                   | -0.14075 | |
| 1   | 0    | 0    | Hypothesized Mean Difference          | 0 | 0 |
| 1   | 0    | 0    | df                                   | 15 | |
| 3   | 0    | 3    | Stat                                 | 1.098 | |
| 0   | 0    | 0    | p(T<|t|) one-tail                       | 0.146 | Accept Null Hypothesis because p > 0.05 (Means are the same) |
| 0   | 0    | 0    | t Critical one-tail                   | 1.753 | Accept Null Hypothesis because p > 0.05 (Means are the same) |
| 0   | 0    | 0    | p(T<|t|) two-tail                       | 0.289 | Accept Null Hypothesis because p > 0.05 (Means are the same) |
| 0   | 0    | 0    | t Critical Two-tail                   | 2.131 | |
| 0   | 0    | 0    | (The two values of p are greater than 0.05 so the Null Hypothesis is accepted) |           | |
| 0   | 0    | 0    |                                        |           | |
Figure 5. Statistical Result of the Number of Errors in Task2

Figure 6. Statistical Result of the Number of Errors in Task3

• Statistical results of the effective of the two websites:

Figure 7. Statistical Results of Number of Clicks in Task1
Figure 8. Statistical Results of Number of Clicks in Task2

Figure 9. Statistical Results of Number of Clicks in Task3

Figure 10. Effective of the two Websites Summary and Completed Task2
2.8. Discussion and Conclusions

On the basis of statistical data of number of clicks in task1, ford is revealing a better effectiveness comparison with BMW. However, the two tested websites are same in terms of efficiency, effectiveness and user satisfaction. Null hypothesis (0a, 0b, 0c) on the basis of three different conducted t-tests is accepted except in the statistical data mentioned above. After evaluating the entire 16 users' responses, in one of the question, Ford got more scores with the statistical significance in comparison to BMW on the ground of “easily identifying of links”.

Three type of errors occurred in the typing of the postcode during the first task in the both websites. On comparing the responses of users, experimenter finds few conclusions regarding the two tested websites:

- Firstly, in an open question, most of the users find frustrated in completing the second task, in the case of BMW. But they are positive in finding the result while surfing the Ford website. Secondly, the BMW website is eye catching and interesting as it possess variety of colors and attractive slideshow. On the other hand, Ford does not find more appealing on the aforesaid grounds. Thirdly, while surfing Ford website, most of the users complained about the small font size and their grey color theme. On the contrast, BMW colors and theme are recognized positively by most of the users. Lastly, the experimenter found few errors in different pages of both the websites. In the case of BMW:-
• On clicking the BMW logo, it’s not redirecting to the homepage as it always happens in normal scenario. For example, in the “find the dealer” page, the Search indicator (►) is not working at all.
In the case of Ford:
• At the bottom of the page, few links like “Focus” in the first column are not redirecting to desired page.

2.9. Limitation and Recommendations
The limitation in this experiment can be:
• The limited number of evaluated criteria.
• The learning effect on the efficiency criteria because of the use of “within users” approach.
• Struggling in a part of the selected tasks in one website can affect the users’ attitude toward the next website.
• The recommendations suggested by the users in the multiple choice questions and in open question as well can be described as below. The experimenter also observes and faces the same problem during the surfing.
• The BMW logo should be navigated to the home page.
• In BMW, the linking of pages to reach a proper model can be improved as user gets confuse while choosing the model in a series.
• In BMW, the menu items should be at the top of the page instead at the bottom to make it more usable.
• In Ford, grey theme could be changed to a bright color as it may turn to more appealing. Color blind people may face difficulty while surfing this website.
• In Ford, the menu style seems to be more traditional according to the 21st century.

3. Appendices

3.1. Tasks’ Sheets

Task1: finding a dealer

A: Initial stage: open http://www.bmw.co.uk  
Script: please do not use search engine.
Goal state: find the nearest dealer for the post code M5 4WT  
Additional criteria: task completion<=5 minutes, total click<=3  
B: Initial stage: open http://www.ford.co.uk  
Script: please do not use search engine.
Goal state: find the nearest dealer for the post code M5 4WT  
Additional criteria: task completion<=5 minutes, total click<=2

Task2: finding a front lights

A: Initial stage: Browser located at http://www.bmw.co.uk  
Script: please do not use search engine.
Second stage: find the 116i sport model.
Goal state: find out if this model has the Front Fog Light as a standard  
Additional criteria: task completion<=7 minutes, total click<=4  
B: Initial stage: Browser located at http://www.ford.co.uk  
Script: please do not use search engine.
Second stage: find the New Focus Titanium X model.
Goal state: find out if this model has the Front Fog Light as a standard
Additional criteria: task completion<=6 minutes, total click<=3

Task3: Website Appearance
A: Initial stage: Browser located at http://www.bmw.co.uk
Script: please look at the website in general and in details by providing your likes and dislike about it, please begin describing by talking.
Second stage: check the website’s screen layout, text, colors, and pictures.
Goal state: find out if the website is more liked than disliked and readable text or not, attractive layout or not and so on.
Additional criteria: task completion<=6 minutes, total click<=5
B: Initial stage: Browser located at http://www.ford.co.uk
Script: please look at the website in general and in details by providing your likes and dislike about it, please begin describing by talking.
Second stage: check the website’s screen layout, text, colors, and pictures.
Goal state: find out if the website is more liked than disliked and readable text or not, attractive layout or not and so on.
Additional criteria: task completion<=6 minutes, total click<=5

3.2. Questionnaire
3.2.1. Demographic Questionnaires
About you:
- Eyesight: normal/corrected
- Occupation: Job description:
- Age: under 20 20-29 30-39 40 or over
- Gender: female/ male
- Education level: high school college bachelor’s degree graduate work post-graduate

Your experience:
- How long have you used computers?
Under 1 year 1-3 years more than 3 years
- Which operating system has you used regularly?
DOS Windows Mac Unix Other
- Which browser has you used regularly?
Mozilla Firefox Google Chrome Internet Explorer Other
- Do you currently drive a car?
Yes No
- When you see a new model car and like it, do you search features of that car to get more information?
Yes No
- How you search the features of the cars?
In the internet dealers store from
- How much are exterior features important for you?
1 2 3 4 5

3.2.2. Users’ Subjective Opinions
Task 1 questions: (www.ford.co.uk)
- Have you ever search any dealer for any car before?
Yes, many times Yes, just once No, never
- How many times did you browse until you find the dealer?
1-3 4-6 more than 6
Do you think if you go out and ask someone about the position of nearest dealer is better than using Ford website? Why?

Is it easy for you to find dealer locator from the main page?

Yes, completely    Yes, Partially    No

It is better to mention about the distance after the nearest dealer search.

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

It is more useful to see the link of the nearest dealer’s official website after the search.

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

Which website is best to find the dealer? Why?

BMW    Ford

Do you think if you go out and ask someone about the position of nearest dealer is better than using BMW website? Why?

Task 1 questions: (www.bmw.co.uk)

How many times did you browse until you find the dealer?

1-3    4-6    more than 6

Is it easy for you to find dealer from the main page?

Yes, completely    Yes, partially    No

It is better to mention about the distance after the nearest dealer search.

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

It is more useful to see the link of the nearest dealers’ official website after the search.

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

Task 2 questions: (www.ford.co.uk)

Have you ever search any front fog lights for any model before?

Yes    No

Did you find the information you are looking for in the Ford website?

Yes    No

Did you know any features of New Focus Titanium X models’ before?

Yes    No

It is easy to find the New Focus Titanium X model features information.

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

It is easy to find the New Focus Titanium X model has front fog lights.

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

It is difficult to complete this task for Ford website?

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

If it is difficult, what could you advice to the developer to navigate it easier?

How many times did you browse until you find feature of front fog light for New Focus Titanium X model?

4.6    7-10    more than 10

Task 2 questions: (www.bmw.co.uk)

Did you find the information you are looking for in the BMW website?

Yes    No

Did you know any features of BMW 116i Sport model before?

Yes    No

It is easy to find the BMW 116i Sport model features information.

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

It is easy to find the BMW 116i Sport model has front fog lights.

1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree

Did you know the front fog light belongs of the safety and technology group of features?
Yes                       No
☐☐ It is difficult to complete this task for BMW website?
1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree
If it is difficult, what could you advise to the developer to navigate it easier?
☐☐ How many times did you browse until you find feature of front fog light for BMW 116i Sport model?
4.6     7-10     more than 10
☐☐ Which website is easier than another to find the front fog light?
BMW     Ford

Task 3 questions:
☐☐ The screen layout works well for you?
1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree
☐☐ The writing is clear and easy to understand?
1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree
☐☐ The text and lettering is legible and readable?
1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree
☐☐ The colors are appropriate for the content and messages being presented? 1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree
☐☐ Graphics and pictures are not distracting?
1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree
☐☐ The theme for the site is a good representation for the content?
1=strongly disagree, 2=disagree, 3=agree somewhat, 4=agree or 5=strongly agree
☐☐ What do you particularly like about www.ford.co.uk?
☐☐ And what do you dislike? What could be improved? What is missing? (for Ford website)
☐☐ What do you particularly like about www.bmw.co.uk?
☐☐ What do you dislike? What can be improved? What is missing? (for BMW website)
☐☐ What are the problems faced in each website?
BMW     Ford
☐☐ Which website from these two do you like? Why?
BMW     Ford
☐☐ Which website from these two do you dislike? Why? What do you think to improve it?
BMW     Ford
☐☐ What do you think the purpose of the ford website? BMW website? Is the design matching to the purpose perfectly?
BMW     Ford
☐☐ Did you find any confuse in the websites? Which one?
BMW     Ford
3.3. Task-based Testing Summary

| TASK                                                                 | COMPLETED TASK? | COMMENTS |
|----------------------------------------------------------------------|-----------------|----------|
| Finding the nearest dealer:                                          |                 |          |
| BMW                                                                   |                 |          |
| Ford                                                                  |                 |          |
| Finding out if the model has front fog lights:                        |                 |          |
| -116i sport (BMW)                                                     |                 |          |
| -New Focus Titanium X (Ford)                                          |                 |          |
| General appearance (Readability, Colour, Font, Images etc…):         |                 |          |
| -BMW                                                                  |                 |          |
| -Ford                                                                 |                 |          |
4. References

[1] P. Sherman and W. Quesenbery, 'Engineering the user experience: UX and the Usability Professionals' Association', Interactions 12, 3 (May. 2005), 38-40.

[2] T. T. Hewett, The role of iterative evaluation in designing systems for usability, Proceedings of the Second Conference of the British Computer Society, human computer interaction specialist group on People and computers: designing for usability, p.196-214, December 1986, York, United Kingdom.

[3] A. Auinger, A. Aistleithner, H. Kindermann and A. Holzinger, Conformity with User Expectations on the Web: Are There Cultural Differences for Design Principles? In: Marcus, A. (ed.) Design, User Experience, and Usability. Theory, Methods, Tools and Practice. Lecture Notes in Computer Science LNCS 6769, vol. 6769, pp. 3-12. Springer, Heidelberg, Berlin, New York, 2011.

[4] M. Leavitt and B. Shneiderman, Research-based web design & usability guidelines. Background and Methodology. US Department of Health and Human Services. Washington, 2006.

[5] A. Badre, Shaping web usability: interaction design in context. Indianapolis, IN: Pearson, 2002.

[6] S. Krug, Don’t make me think! 2nd ed. A common sense approach to web usability. Indianapolis: New Riders, 2006.

[7] M. Matera, F. Rizzo and G. Carugi, Web usability: Principles and evaluation methods. In E. Mendes & N. Mosley (Eds.), Web engineering (pp. 143–180). Berlin, Germany: Springer, 2006.

[8] T. Tullis and B. Albert, Measuring the user experience: collecting, analyzing, and presenting usability metrics. Burlington, MA: Elsevier, 2008.

[9] P. Alpar, Satisfaction with a Web Site: Its Measurement, Factors and Correlates. In A.-W. Scheer & M. Nüttgens (Hrsg.), Electronic Business Engineering /4. International Tagung Wirtschaftsinformatik. Heidelberg: Physica Verlag, 1999.

[10] C. Courage and K. Baxter, Understanding Your Users: A Practical Guide to User Requirements. Morgan Kaufmann Publishers, San Francisco, CA, 2005.

[11] A. Dean and D. Voss. Design and analysis of experiments. Springer Verlag, New York, NY, USA, 1999.

[12] D. Martin, Doing psychology experiments 4th ed.[S.I.] Brooks/Cole Pub. Co, Pacific Grove, Calif, 1996.