A Review of Website Measurement for Website Usability Evaluation

Nur Sukinah Aziz, Noor Suhana Sulaiman, Wan Nur Idayu Tun Mohd Hassan, Nur Liyana Zakaria, Azliza Yaacob
Faculty of Computer, Media and Technology Management, University College TATI, Telok Kalong, Kemaman, Terengganu
nursukinah@uctati.edu.my

Abstract. There are many website usability measurement. Therefore, researcher that focus on usability measurement need to select the best method for usability evaluation to get the best result. This paper review website measurement that has classify to usability inspection, usability inquiry and usability testing. Some of usability measurement focus on experts only to measure the evaluation and some of measurement focus on real users. Each of website measurement have advantage and disadvantage. Researcher need to determine the scope, context and the reason when evaluate usability for website. This paper found that survey is common method that used by researcher because it save cost and can gather a population easily.

1. Introduction
Nowadays, websites are used widely all over the world as a medium of communication for information or services. Organizations use websites to market their product and services. Users will always use the website if the website can achieve their task or goal in searching information or using the services more quickly, easily and effectively [1]. Website evaluation determines the quality of the website. The advantages of usability evaluation are related to ergonomics, website quality and quality management evaluation processes. Different usability approaches are evaluated and focus is placed on the advantages of defining usability in terms of meeting user needs by ensuring quality of use [2]. Measurement of usability gives advantages such as can predict, guarantee and enhance website quality. The regulation and enhancement of production processes and deciding on the approval of a website. There are four basic ways to evaluate usability such as automatically (usability measures computed by running a user interface specification through evaluation software), empirically (usability assessed by testing the interface with real users), formally (using exact model and formulas to calculate usability measures) and informally (based on rules of thumb and the general skill, knowledge and experience of evaluators ) [3]. There are many approaches to usability evaluation such as testing, inspection, inquiry, analytical modelling and simulation [4]. All the evaluation methods have one common characteristic that is dependent on user judgement.

2. Usability Testing
Usability lab testing focus on the experience and comments from users that used the websites or in a scenario-based environment [5]. Most usability testing is at the phase of user-centered design
(UCD). Usability lab testing only involves a small group of users [6]. It records the user behaviour and cognitive processes to see what users actually feel and how users use the website. There is a computer lab with a workstation that can record keystrokes using video tapes of the user’s hand movement and spontaneous comments while using the website. During the observation session, all the data will be recorded and the data will be analysed with many functions. The examples of usability testing are co-discovery learning, question-asking protocol, performance measurement, retrospective testing, teaching method, coaching method, remote testing, retrospective testing, shadowing method and many more. Thinking aloud protocol is an approach that encourages the respondents to share his or her opinion while navigating the site. The users will provide the website to be tested and a set of tasks to perform. The users need to perform the tasks of using the website and explain what they are thinking about while doing the tasks. This method allows the researcher to understand the user’s approach in using the website. The researcher can see how the user navigates in the website, where areas of confusion exist and where the users see something unexpected. But it has one disadvantage, it rarely works well and it is not natural for most users to speak aloud. Besides that, eye-tracking is also used in usability testing. The eye-tracking approach allows researchers to see what users see and where they look by tracking their eye movements and the length of time they fixate on a certain part of screen or object on the screen.

3. Usability inspection

Usability inspection is the generic name for a set of methods based on having evaluators inspect or examine usability-related aspects of a user interface [3]. This requires usability specialists or software developers, users and other professionals to examine or judge either the prototype or each element of interface following established usability principles. The examples of usability inspection are cognitive walkthrough, heuristic evaluation, feature inspection, formal usability inspection and many more. Cognitive walkthrough uses a more explicitly detailed procedure to simulate a user’s problem solving process at each step in the human-computer dialogue, checking to see if the simulated user’s goals and memory for actions can be assumed to lead to the next correct action. Heuristic evaluation or expert review verifies the interface or web design that may include site layout and structure, navigation tools, search function, fonts and colour, and give opinion about the pros and cons. The heuristic evaluation involves usability specialists to judge whether the element follows established usability principles or the heuristic. Expert reviews give new, outside perspective to the website and which is the best in web design. But there are limitations using this approach such as the expert review did not understand the intended audience or the company policies and budget constraints. Three (3) experts review involved in the research and it sufficient to gather the information from experts [7]–[9]. The experts need to select that had experience and involvement in website usability more than five (5) years.

Feature inspection can involve not only evaluation of a function but also the design of that function whether its design meets the needs of intended end users. Feature inspection lists the sequences of features used to accomplish typical tasks, checks for long sequences, cumbersome steps, steps that would not be natural for users to try and steps that require extensive knowledge or experience in order to assess a proposed feature set.

Pluralistic walkthrough is an approach that meets with users and developers whereby people step through a scenario discussing usability issues associated with dialogue elements involved in the scenario steps. Standards inspection has an expert on some interface standards and inspects the interface for compliance. Standards are aimed at increasing the degree to which a given interface is in the range of other systems on the market that follow the same standards. Guideline checklist is an
approach where the interface is checked for conformance with a comprehensive list of usability guidelines. This method is fairly rare in practice because it requires a high degree of expertise since it contains an order of 1,000 guidelines. Formal Usability Inspections involve various participants that have well defined responsibilities. Inspections are performed through a six-step process; planning, a kickoff meeting, a preparation phase where inspectors review the interface individually, the main inspection review when the inspectors’ lists of usability problems are merged and follow-up phase where the effectiveness of the inspection process itself is assessed. Consistency Inspection is where the designers representing multiple projects inspect an interface to see whether it does things in a way that is consistent with their own designs. Consistency inspection is aimed at evaluating consistency across the family of products that has been evaluated by an inspection team.

4. Usability inquiry

This involves experts to get information about the user requirement for the system by communicating with them or observing them while users are using the system. There are many types of usability inquiry such as field observation that is visiting the users and observing them work, interviews which are to conduct exploratory studies where one does not know yet what one is looking for, or focus groups where users are brought together to discuss new concepts and identify important issues. Interviews are a good means to gain deep data from a small number of respondents. Besides that, surveys can understand the preferences of a broad base of users about an existing or potential website or system. It can use larger samples to generalize an entire population. Ethnographic study involves observing users in the place where they would normally use the website or system such as at work, home, coffee bar or other places to gather data. Other usability inquiries are proactive field study, logging actual use, proactive field study, contextual inquiry, journal sessions, self-reporting logs and screen snapshots.

5. Survey

A survey commonly is to collect data representative of a population [10]. The survey methodology is suitable when asking about users perception, feelings, opinions and thoughts. Surveys also collects data relating to beliefs, attitudes and motives [11]. It means that survey is assessing information about the sample and enables the researcher to draw conclusions on generalizing the findings from a sample of responses to a population [12]. The survey method is quick, cost effective or inexpensive, can gather a lot of data and can be administered on a large population [13]. Therefore, this method is suitable for a research with a large sample size. The main survey instrument is a questionnaire. Before distributing the questionnaire to the actual study, there are a few steps to follow to evaluate the questionnaire, such as content analysis, pre-test and pilot test [14],[15]. Those steps are to check the questions if the questionnaire are understandable by respondents, well written and to improve the questionnaire [16]. It is also to evaluate the reliability and validity of the instrument [17], [18].

6. Questionnaires

Questionnaires are the most frequently used tool for usability evaluation in a survey. It is easier and cheaper than any other approach to gather user feedback [19]. There are three types of questionnaire, pre-test questionnaire, post-task questionnaire and post-test questionnaire. A questionnaire is a structured list of questions used to elicit information from respondents in freestyle writing or by ticking off pre-defined answers [20]. A pre-test questionnaire is to get more information about participants before they start working with the product. A post-task questionnaire is to get immediate feedback after each scenario. A post-test questionnaire is to get feedback about the whole experience. Many users are involved in this evaluation. There are many types of standard post-test questionnaires based on the specific field and criteria such as System Usability Scale (SUS), Questionnaire for User Interface Satisfaction (QUIS) [21], Computer System Usability Questionnaire (CSUQ), Words (adapted from Microsoft’s Product Reaction Cards) [22] and others. All these questionnaires are widely used and readily available. Basically, questionnaires are evaluations that access the perception
from the user’s point of view of the websites. Based on a study by [22], they found that SUS and CSUQ give the most consistent results with varying sample sizes. In this study, they used five questionnaires on two websites to determine user preference [22], [23]. However, a questionnaire is not always directly applicable depending on the application domain and may not cover all aspects in need of evaluation [24].

i. System Usability Scale (SUS) - 1986

System Usability Scale (SUS) was developed by John Brooke at Digital Equipment Corporation in 1986. It is also known as a “quick and dirty” survey scale to evaluate the usability that allows the usability practitioner to quickly and easily evaluate. It uses a 10 item questionnaire with the response based on a 5-point scale. The SUS questionnaire has its own method to calculate the score. The even-numbered items are positive statements and the odd-numbered items are negative statements to balance the responses. We can add other questions if we need to get more information about our study. But the alternative questions cannot be included in the SUS score [23]. To calculate the SUS score, first get the sum of the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1, 3, 5, 7 and 9, the score contribution is the scale position minus 1. For items 2, 4, 6, 8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall SUS score. SUS scores have a range of 0 to 100 [25].

ii. Computer System Usability Questionnaire (CSUQ) - 1995

Developed by James Lewis at IBM. It uses 19 questions on a 7-point scale of “Strongly Disagree” to “Strongly Agree”, including “Not Applicable” (N/A). CSUQ is very similar with Lewis’s Post-Study System Usability Questionnaire (PSSUQ) with only minor changes in wording. CSUQ is different from SUS because all statements in CSUQ is worded positively. CSUQ may be viewed in four main categories: System Usefulness, Information Quality, Interface Quality and Overall Satisfaction (Tullis & Albert, 2013). CSUQ is suitable for usability study in a non-laboratory setting.

iii. Website analysis and measurement inventory (WAMMI) - 1999

WAMMI was developed by the Human Factors Research Group (HFRG), University College Cork in 1999 (Tullis & Albert, 2013). WAMMI is an evaluation tool for websites. It is based on a questionnaire filled by visitors of a website, and gives a measure of how useful and easy it is to use the visitors found about the site. The WAMMI report provides the following information such as overall usability score and the general rating of a website, detailed usability profile in terms of five usability scales: attractiveness, control, efficiency, helpfulness, and learnability. WAMMI has 20 items in the questionnaire that uses the 5-point scales. Some of the statements are positive and others have negative wording. The advantage of WAMMI is that additional questions can be added and the associated rating scale in WAMMI has already been used in the evaluation of many websites. The results are in the form of comparison against their reference database built from tests of hundreds of websites.

Table 2.12 shows the questionnaires that have been tested empirically based on previous research in usability and WU. The table shows the acronym of the questionnaire, instrument, author or reference, institution, number of item in the questionnaire and the detail or attributes in the questionnaire.

| Acronym | Instrument | Author / Reference | Institution | Item | Detail / Attributes |
|---------|------------|--------------------|-------------|------|--------------------|
| CUSI    | Computer Usability | [27]                | Human Factors | 22 questionnaires | Affect (the degree to |
| Acronym   | Instrument                                      | Author / Reference | Institution          | Item                                | Detail / Attributes                                                                 |
|----------|-------------------------------------------------|--------------------|----------------------|------------------------------------|-------------------------------------------------------------------------------------|
|          | Satisfaction Inventory                          | Research Group (HFRG), University College Cork, | re which users like the computer system) |
|          |                                                 | Chin et al. (1988)/[28] | University of Maryland | 27 item questionnaire using 10 point scale |
|           | QUIS Questionnaire for User Interface Satisfaction |                                 |                      |                                    |
|          | SUS System Usability Scale                      | [25] [23]           | Digital Equipment Corporation | It uses a 10 item questionnaire with response based on a 5-point scale. The SUS questionnaire has its own method to calculate the score |
|          | PUEU Perceived Usefulness and Ease of Use       | [29]/[28]           | IBM                  | 12 item questionnaire with 7 likert scale |
|          | NAU Nielsen’s Attributes of Usability            | Nielsen (1993a)/[28] | Bellcore             |                                    |
|          | CSUQ Computer System Usability Questionnaire    | [30]                | IBM                  | 19 item questionnaire using 7 scale |
|          | ASQ After Scenario                              | [30]                | IBM                  | 3 item questionnaire               |

- Competence (the degree to which users feel supported by the computer system).
- Screen
- Terminology and System Information
- Learning
- System capabilities
- Overall reaction to the software
- Perceived Usefulness
- Ease of Use
- Learnability
- Efficiency
- Memorability
- Errors
- Satisfaction
- Overall
- System Usefulness
- Information Quality
- Interface Quality
- Ease of task completion
| Acronym | Instrument | Author / Reference | Institution | Item | Detail / Attributes |
|---------|------------|-------------------|-------------|------|--------------------|
| PSSUQ   | Poststudy System Usability Questionnaire | [30] [31] IBM | 19 item questionnaire using 7 point scales | • Time to complete a task • Adequacy of support information |
| PUTQ    | Purdue Usability Testing Questionnaire | Lin et al. (1997) /[32] Purdue | 8 factor of Human-Computer in terms of software usability |
| SUMI    | Software Usability Measurement Inventory | Kirakowski & Corbett (1994) Kirakowski (1996) Human Factors Research Group (HFRG), University College Cork | 50 item questionnaire using 3 likert scale | • Efficiency • Affect • Helpfulness • Control • Learnability |
| MUMMS   | Measuring the Usability of Multi-Media Software | /[32] | MUMMS are used to evaluate multimedia software |
| WAMMI   | Website Analysis And Measurement Inventory | Human Factors Research Group (HFRG), University College Cork | 20 item questionnaire using 5 point scale | • Attractiveness • Control • Efficiency • Helpfulness • Learnability |

7. Summary of website measurement
Many aspects which took into consideration the importance in usability evaluations. Anything that surrounds it is strongly affected by the user experience. Table 7.1 is a summary about website measurement and the advantages and disadvantages of each category.
| Categories | Method                | Advantages                                                                 | Disadvantage                                                                 |
|------------|-----------------------|----------------------------------------------------------------------------|----------------------------------------------------------------------------|
| Testing    | Thinking Aloud        | - Able to find why the problems occur                                     | - Time consuming                                                           |
|            |                       | - Small number of test users                                               | - Can be expensive to hire participants                                    |
|            |                       | - Low time in relation to other evaluation methods                         | - The users may feel uncomfortable                                          |
|            |                       | - Direct interaction of the users with the transactional web application   |                                                                           |
|            |                       | - Can get direct feedback from real potential users                        |                                                                           |
| Testing    | Formal Evaluation     | - Objective method                                                         | - Time consuming                                                           |
|            |                       | - Provide substantive depth in quantitative data                           |                                                                           |
| Testing    | Query Technique       | - Provide qualitative and quantitative data                                | - Time consuming                                                           |
|            |                       | - Simple and cheap                                                         |                                                                           |
| Inspection | Prototype             | - Issues in design can be identified                                       | - Time consuming to create                                                 |
|            |                       | - Complete functionality can be tested                                     | - More expensive to develop                                               |
| Inspection | Heuristic Evaluation  | - Easy to perform; cheap                                                   | - Focus on problems                                                        |
|            |                       | - No planning required                                                     |                                                                           |
|            |                       | - Able to find many problems (both major and minor problems)              |                                                                           |
|            |                       | - Not time-consuming (no users involved)                                   |                                                                           |
| Inspection | Cognitive walkthrough | - Get the opinion from expert                                              | - May be expensive to pay the experts                                      |
|            |                       | - Puts focus on the user                                                   | - Need more than 1 expert                                                 |
|            |                       | - Recognition of user’s goal                                               | - Slow because have many phases                                           |
| Inspection | Pluralistic walkthrough| - Can provide early performance and satisfaction data from users when a   |                                                                           |
|            |                       |   user interface prototype is not available.                               |                                                                           |
| Inquiry    | Individual interview  | - Enable to learn about things that cannot be directly observed           | - Time consuming                                                           |
|            |                       | - Structured data can be collected                                         | - Expensive                                                               |
|            |                       | - Planning can enable in-depth discussion                                   |                                                                           |
| Inquiry    | Focus group           | - Can provide speedy results                                              | - Recruitment can be expensive                                             |
|            |                       | - Structured data can be collected                                         | - Time consuming                                                           |
|            |                       | - Planning can enable in-depth discussion                                   | - Controlled settings may affect behaviours                                |
|            |                       | - Simple to administer                                                     | - Data can be difficult to analyse                                         |
8. Conclusion

There are many methods in website measurement. Researcher need to determine a few issues in selecting the best method for website measurement such as cost, time and respondent. Usually, researchers select survey because of it quick and cost effective. Data collection could be executed in several ways such as online survey, email survey and telephonic survey. Self-administered questionnaire refers to ‘a data collection technique in which the respondents reads the survey questions and records his or her responses without the presence of a trained interviewer’. The questionnaires were handed to the respondents and it could create a rapport with the respondent. Thus, it could motivate the respondent to answer the questions listed in the questionnaires. Besides it was the cheaper way to collect the data, it also will lead to the better rate of response rate. Therefore an understanding of the context's impact and how it affects the usability assessment process is required. In exploring the role of context in usability assessments, this research has provided the as is for our future work. The line of investigation needed to move forward in developing a context-encompassing usability evaluation system has also been developed.

Acknowledgement
This research has been funded with supported by short term (STG) grant from University College TATI (UC TATI), Malaysia, Grant 1/2018 (GPJP 2018), No. Grant 9001-1801. The authors fully acknowledged University College TATI (UC TATI), Malaysia for the approved fund which makes this important research viable and effective. The authors gratefully acknowledge support from all research partners.

References

[1] S. Giraud, P. Thérouanne, and D. D. Steiner, “Web accessibility: Filtering redundant and irrelevant information improves website usability for blind users,” Int. J. Hum. Comput. Stud., vol. 111, no. April 2017, pp. 23–35, 2018, doi: 10.1016/j.ijhcs.2017.10.011.

[2] N. S. Aziz and A. Kamaludin, “Development of Website Usability Instrument Based on Experts Review,” 2015 2nd Int. Conf. Commun. Comput. Eng., vol. 11, no. 1984, pp. 1–10,
2015.
[3] J. Nielsen and R. L. Mack, *Usability Inspection Methods*. John Wiley & Sons, Inc, 1994.
[4] J. Wang and S. Senecal, “Measuring Perceived Website Usability,” *J. Internet Commer.*, vol. 6, no. 4, pp. 97–112, 2008.
[5] S. Batra and R. R. Bishu, “Web usability and evaluation: Issues and concerns,” *Usability Int. HCI Cult.*, pp. 243–249, 2007.
[6] A. Ai-wabil and R. Ai-khalifa, “A Framework for Integrating Usability Evaluations Methods: The Mawhiba Web Portal Case Study,” *Methodology*, 2009.
[7] L. Crocker, A. Maria Llabre, and M. D. Miller, “The Generalizability of Content Validity Ratings,” *J. Educ. Meas.*, vol. 25, no. 4, pp. 287–299, 1988.
[8] S. N. Haynes, D. C. S. Richard, and E. S. Kubany, “Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods Introduction to Content Validity,” *Psychol. Assess.*, vol. 7, no. 3, pp. 238–247, 2012.
[9] E. Delgado-rico, H. Carretero-dios, and W. Ruch, “Content validity evidences in test development;,” *Int. J. Clin. Heal. Psychol.*, vol. 12, pp. 449–459, 2012.
[10] C. W. Craighead, D. J. Ketchen, K. S. Dunn, and G. T. M. Hult, “Addressing Common Method Variance: Guidelines for Survey Research on Information Technology, Operations, and Supply Chain Management,” *IEEE Trans. Eng.*, vol. 58, no. 3, pp. 578–588, 2011.
[11] S. O. Ogara, C. E. Koh, and V. R. Prybutok, “Investigating factors affecting social presence and user satisfaction with Mobile Instant Messaging,” *Comput. Human Behav.*, vol. 36, pp. 453–459, 2014, doi: 10.1016/j.chb.2014.03.064.
[12] I. M. Saidon, “Moral Disengagement in Manufacturing: A Malaysian Study of Antecedents and Outcomes Intan Marzita Saidon,” Curtin University, 2012.
[13] K. I. Al-Qeisi, “Analyzing the Use of UTAUT Model in Explaining an Online Behaviour: Internet Banking Adoption,” Brunel University, 2009.
[14] J. Lazar, *Web Usability: A User-Centered Design Approach*. Pearson Education, 2006.
[15] B. Kitchenham and S. L. Pfleeger, “Principles of survey research part 4: questionnaire evaluation,” in *ACM SIGSOFT Software Engineering Notes*, 2002, vol. 27, no. 3, p. 20, doi: 10.1145/638574.638580.
[16] S. Presser et al., “Methods for Testing and Evaluating Survey Questions,” *Public Opin. Q.*, vol. 68, no. 1, pp. 109–130, 2004, doi: 10.1093/poq/npn001.
[17] S. E. Fawcett, M. a. Waller, J. W. Miller, M. a. Schwieterman, B. T. Hazen, and R. E. Overstreet, “A trail guide to publishing success: Tips on writing influential conceptual, qualitative, and survey research,” *J. Bus. Logist.*, vol. 35, no. 1, pp. 1–16, 2014, doi: 10.1111/jbl.12039.
[18] K. Salleh, S. C. Chong, S. N. Syed Ahmad, and S. O. S. Syed Ikhsan, “Learning and knowledge transfer performance among public sector accountants: an empirical survey,” *Knowl. Manag. Res. Pract.*, vol. 10, pp. 164–174, Jun. 2012, doi: 10.1057/kmrp.2011.46.
[19] S. Elling, L. Lentz, and M. de Jong, “Website Evaluation Questionnaire: Development of a Research-Based Tool,” in *International Conference on Electronic Government*, 2007, pp. 293–304.
[20] N. O. Bernsen and L. Dybkjaer, *Multimodal Usability*. Springer, 2010.
[21] A. Madan and S. Kumar Dubey, “Usability evaluation methods: A literature review,” *Int. J. Engng. Sci. Technol.*, vol. 4, no. 2, pp. 590–599, 2012.
[22] T. S. Tullis and J. N. Stetson, “A Comparison of Questionnaires for Assessing Website Usability,” pp. 1–12, 2004.
[23] C. M. Barnum, *Usability Testing Essentials: Ready, Set...Test!* Morgan Kaufmann, 2011.
[24] R. Tezza, A. C. Bornia, and D. F. de Andrade, “Measuring web usability using item response theory: Principles, features and opportunities,” *Interact. Comput.*, vol. 23, pp. 167–175, 2011, doi: 10.1016/j.intcom.2011.02.004.
[25] J. Brooke, “SUS - A quick and dirty usability scale,” *Usability Eval. Ind.*, vol. 189, no. 194, pp.
4–7, 1996.

[26] T. Tullis and B. Albert, *Measuring the user experience collecting, analyzing, and presenting usability metrics*, 2nd ed. Morgan Kaufmann, 2013.

[27] J. Kirakowski and A. Dillon, “The computer user satisfaction inventory (CUSI): Manual and scoring key,” 1988.

[28] M. De Marsico and S. Levialdi, “Evaluating web sites: exploiting user’s expectations,” *Int. J. Hum. Comput. Stud.*, vol. 60, no. 3, pp. 381–416, Mar. 2004, doi: 10.1016/j.ijhcs.2003.10.008.

[29] F. D. Davis, “Perceived Usefulness, Perceived Ease Of Use, And User Acceptance,” *MIS Q.*, vol. 13, no. 3, pp. 319–339, 1989, doi: 10.2307/249008.

[30] J. R. Lewis, “IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use,” *Int. J. Hum. Comput. Interact.*, vol. 7, no. 1, pp. 57–78, 1995, doi: 10.1080/10447319509526110.

[31] A. Bangor, P. T. Kortum, and J. T. Miller, “An Empirical Evaluation of the System Usability Scale,” *Int. J. Hum. Comput. Interact.*, vol. 24, no. 6, pp. 574–594, Jul. 2008, doi: 10.1080/10447310802205776.

[32] J. A. Hyman, “Towards an Understanding of Mobile Website Contextual Usability and its Impact on Mobile Commerce by,” 2012.